User Manual

An implementation of the model described in the main manuscript is provided as an R script named "model.R". The script contains functions for the estimation of haplotype frequencies, multiplicity of infection (MOI), and two linkage disequilibrium (LD) measures, i.e., D' and R^2, from a pair of multi-allelic loci. Additionally, the code in the script "tutorial.R" serves as template for deriving the estimates from empirical data.

The scripts and some example datasets can be found on GitHub (https://github.com/Maths-against-Malaria/generalModel).

Loading the R script

Suppose the main R script is stored in a directory "<PATH>/STRModel.R". First, the script has to be loaded in the R environment (e.g., R Studio, VS Code, or an R terminal) using the following code:

```
# Load external resources
    source("/Users/christian/Documents/phd/models/generalModel/src/model.R")
```

Here, we assume that the <PATH> containing the R script is "/home/johndoe/Documents/".

Importing data

A dataset needs to be imported first (see below for the required format). Assume the dataset "example_dataset1.xlsx" (provided with the script) is downloaded and stored in the folder "Documents" whose path is given by "/home/johndoe/Documents/". The dataset can be imported using the R package "openxlsx". However, the package is not present by default and might need to be installed using the following code:

```
# Install library "openxlsx"
    #install.packages("openxlsx")
```

The library is then loaded upon successful installation and the data imported using the code:

```
# Load library "openxlsx"
library(openxlsx)
```

A more comprehensive documentation exists as a guide to use the package "openxlsx". Note that other packages can be used to import ".xlsx" files in R and represent a good alternative to the "openxlsx" package described above. Moreover, functions such as "read.csv" and "read.table" can be used to import data in the ".csv" and ".txt" formats, respectively.

Standard input format and data transformation

The methods are designed for data containing information from a pair of multi-allelic markers, e.g., microsatellites markers with n_1 and n_2 alleles, respectively. First, the desired data format is explained. Second, it is explained how custom data can be converted into this format.

For each record (corresponding to one sample) the data indicates the absence and presence of the alleles found at both molecular markers, in the following convention. At marker k, the absence and presence of the alleles correspond to a 0-1 vector of length n_k ($k = 1, 2$). This corresponds to a binary number between 0 and $2^{n_k} - 1$, where the vector $\mathbf{0} = (0, \ldots, 0)$ corresponds to missing data, $(1, 0, \ldots, 0)$ indicates the presence of the first allele, and $(1, 1, \ldots, 1)$ indicates the presence of all alleles.

A dataset of sample size N is an array with N rows in which the entries are numbers from 0 to $2^{n_1} - 1$ and 0 to $2^{n_2} - 1$ at the first and second marker, respectively. As an example, the following dataset of sample size $N = 100$, with $n_1 = 2$ and $n_2 = 3$ alleles at the first and second marker has the correct format:

Absence/presence of alleles at the first marker are encoded by numbers from 0 to $2^2 - 1 = 3$, and by numbers from 0 to $2^3 - 1 = 7$ at the second marker. For the second record (ID2) the number 2 at the first marker corresponds to the 0-1 vector (0,1), indicating the absence of the first and presence of the second allele, while

ID	Marker1	Marker2
ID1	3	7
ID2	2	5
ID3	1	2
⋮	⋮	⋮
ID99	0	0
ID100	0	3

the entry 5 at marker 2 corresponds to the 0-1-vector (1,0,1) that represents the presence of only the first and third allele. For the last record (ID100), entry 0 at marker 1 corresponds to the vector (0,0) indicating missing data, while entry 3 is equivalent to the 0-1 vector (1,1,0), indicating the presence of the first and second alleles.

This format is referred to as ***standard input format***. The ID column in the dataset is optional and can be omitted.

If data is not present in the format outlined above, it needs to be transformed. This can be done easily if the data is already in the following 'more natural' format, for which each record (sample) is represented by multiple **consecutive** rows, which list the alleles found at the respective markers. The first column has to contain the sample ID.

Assume a dataset with four molecular markers and the STR alleles (corresponding to distinct sequence lengths): (i) 130, 133 at marker 1; (ii) 201, 207, 210 at marker 2; (iii) 89, 91, 94, 99 at marker 3; and (iv) 140, 145, 148 at marker 4. Assume the following structure:

ID	M1	M2	M3	M4
ID1	130	207	99	140
ID1	133	201		
ID1		210		
ID2	133	201	99	140
ID2	133	210	91	145
ID3	130	207	91	148
⋮	⋮	⋮	⋮	⋮
ID99			89	145
ID99			99	148
ID100		201	94	148
ID100		207		140

Consider sample ID1. The two alleles 130 and 133 correspond to the 0-1 vector (1,1) at marker 1, the three alleles 201, 207, and 210, to (1,1,1) at marker 2, the allele 99 to (1,0,0,0) at marker 3, and the allele 140 to (1,0,0) at marker 4.

The function "data_format(, id=TRUE)" takes the data in this 'more natural' format and creates a list as an output, with the data transformed into the standard input format as the first element. The second element is a list, which contains a string vector with the alleles occurring at each marker. The third element contains a vector with the number of alleles found at each marker. The optional boolean argument "id" (default id=TRUE) indicates whether the transformed dataset in the standard input format should contain the sample ID in the first column. The following code first imports the data "/home/johndoe/Documents/example_dataset4.xlsx" into "DATA1", transforms it into the standard input format, and saves the output list as "Ex.data":

```
## Import the dataset
datasetNaturalFormat <- read.xlsx('/Users/christian/Documents/phd/models/generalModel/exampleDatasets/e

# Transform the data to the standard format
```

```
datasetStandard <- convertDatasetToStandardFormat(datasetNaturalFormat, 2:ncol(datasetNaturalFormat))
```

```
##  $ Ex.data
##  [[1]]
##  ID    M1   M2   M3   M4
##  ID1    2    4    4    4
##  ID2    2    4    2    1
##  ID3    1    4    2    1
##  ...    ...  ...  ...  ...
##  ID98   2    4    2    1
##  ID99   3    4    4    5
##  ID100  1    5    4    4

##  [[2]]
##  [[2]]$M1
##  [1] "130" "133"

##  [[2]]$M2
##  [1] "201" "207" "210"

##  [[2]]$M3
##  [1] "89" "91" "94" "99"

##  [[2]]$M4
##  [1] "140" "145" "148"

##  [[3]]
##  M1 M2 M3 M4
##   2  3  4  3
```

The first element of the above list is in the desired standard input format but contains more than 2 markers.

If data is given in a different format, the R-package **MLMOI** provides a flexible function to import the data into this format. This package also helps to detect data entry errors. To install the package, load it and access the documentation type

```
# Install library "MLMOI"
#install.packages("MLMOI")

# Load library
#library("MLMOI")

# Consult documentation
?MLMOI
```

```
## No documentation for 'MLMOI' in specified packages and libraries:
## you could try '??MLMOI'
```

Haplotype frequencies and MOI estimates

Assume that the script "STRModel.R" and the data set <DATA> were loaded (see above).

The function "mle(<DATA>, <n1n2>,...)" derives the maximum likelihood estimates (MLEs) of the 2-marker haplotype frequencies and the MOI parameter. The arguments are the dataset in standard input format (<DATA>), and a vector containing the number of alleles at the first and second marker (<n1n2>) and several optional arguments.

If <DATA> does not contain sample IDs in the first column, the argument "id = FALSE" must be specified.

The output of the function "mle(<DATA>, <n1n2>,...)" is a list with three elements: (i) the MOI parameter $\hat{\lambda}$; (ii) the non-zero haplotype frequencies $\hat{p}$; and (iii) a matrix of all detected haplotypes with estimated non-vanishing frequency. Alleles at marker k are denoted by the numbers $0, \ldots, n_k - 1$ $(k = 1, 2)$.

```
## Choose markers of interests
markers <- 1:4
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
## [1] 0.968123
##
## $haplotypes_frequencies
##                [,1]
## 1112  2.094589e-02
## 1121  5.487696e-02
## 1123  9.939004e-03
## 1131  1.305780e-02
## 1133  1.463578e-02
## 1221  2.980299e-02
## 1243  1.003814e-02
## 1311  7.916535e-03
## 1312  2.665850e-02
## 1321  1.322501e-01
## 1341  2.111656e-02
## 1323  2.726089e-02
## 1331  1.379620e-02
## 1333  2.472659e-02
## 1111 3.123175e-183
## 1122 2.845724e-151
## 1322 1.302211e-193
## 2121  4.323393e-02
## 2111  9.075848e-03
## 2133  3.600153e-02
## 2131  2.292512e-56
## 2123  7.210104e-90
## 2221  5.602034e-02
## 2222  1.574441e-02
## 2232  1.574441e-02
## 2231  1.338446e-02
## 2233  2.700632e-15
## 2311  2.610855e-02
## 2312  1.574441e-02
## 2313  2.216986e-02
## 2321  1.158664e-01
## 2333  5.001950e-02
## 2331  1.110850e-02
## 2323  9.446176e-03
## 2343  2.435875e-02
## 2141  1.857786e-02
## 2341  1.666925e-02
## 2143  9.613822e-03
## 2241  1.961211e-02
## 2211  3.051500e-08
```

```
## 2113  9.951688e-03
## 2213  5.224381e-57
## 2223  9.096933e-19
## 1113  7.842207e-03
## 1141  1.019287e-02
## 1222  7.842207e-03
## 1212  7.842207e-03
## 2212 8.151060e-203
## 1232 8.151060e-203
## 1231  3.751307e-61
## 1313  1.199195e-02
## 1343  2.239358e-49
## 1241  3.787025e-54
## 1223  8.814834e-03
## 1233  6.897002e-68
## 1143  2.288823e-70
## 2243  7.179864e-65
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
## [16,]    1    1    2    2
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
```

```
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
## [48,]    2    2    1    2
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

As an additional example, consider the list "Ex.data". The data set in standard input format is contained as first element, but contains information from 4 markers. Suppose the estimates are desired for data from markers 3 and 4. The following code provides the estimates:

```
### Selecting the data without the column for sample IDs:
#data <- Ex.data[[1]][,c(4,5)]
### Selecting the number of alleles at markers 3 and 4:
#GA <- Ex.data[[3]][c(3,4)]
### Estimating the MLEs
#mle(data, GA, id = FALSE)
```

Haplotype frequencies using a plug-in estimate for MOI

The function "mle(<DATA>, <n1n2>,...)" allows to derive the MLEs for haplotype frequencies using a plug-in value for the MOI parameter (e.g., it has been independently estimated), i.e., the function provides the profile-likelihood estimates for haplotype frequencies given a fixed value for the MOI parameter. The argument "plugin=$\hat{\lambda}_{\text{plugin}}$" specifies that $\hat{\lambda}_{\text{plugin}}$ should be used as plug-in estimate for the MOI parameter.

Consider the data "example_dataset1.xlsx" (with $n_1 = 2$ and $n_2 = 3$ allels for marker 1 and 2, respectively). Assuming the plug-in $\hat{\lambda}_{\text{plugin}} = 0.2$ for the MOI parameter, MLEs for the haplotype frequencies are obtained by running the code:

```
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
## [1] 1
##
## $haplotypes_frequencies
```

```
##              [,1]
## 1112  2.092049e-02
## 1121  5.489177e-02
## 1123  9.927741e-03
## 1131  1.304207e-02
## 1133  1.461721e-02
## 1221  2.980708e-02
## 1243  1.002225e-02
## 1311  7.907381e-03
## 1312  2.663251e-02
## 1321  1.323901e-01
## 1341  2.112388e-02
## 1323  2.726499e-02
## 1331  1.378108e-02
## 1333  2.472385e-02
## 1111 5.055437e-182
## 1122 9.793412e-151
## 1322 2.084481e-192
## 2121  4.326242e-02
## 2111  9.066195e-03
## 2133  3.596598e-02
## 2131  3.727526e-56
## 2123  1.274991e-89
## 2221  5.601583e-02
## 2222  1.572357e-02
## 2232  1.572357e-02
## 2231  1.337858e-02
## 2233  3.350377e-15
## 2311  2.608186e-02
## 2312  1.572357e-02
## 2313  2.215044e-02
## 2321  1.161752e-01
## 2333  4.999194e-02
## 2331  1.110659e-02
## 2323  9.450508e-03
## 2343  2.432779e-02
## 2141  1.855058e-02
## 2341  1.667837e-02
## 2143  9.598406e-03
## 2241  1.958559e-02
## 2211  3.386660e-08
## 2113  9.937696e-03
## 2213  5.132737e-57
## 2223  1.170623e-18
## 1113  7.830881e-03
## 1141  1.018318e-02
## 1222  7.830881e-03
## 1212  7.830881e-03
## 2212 1.371895e-201
## 1232 1.371895e-201
## 1231  4.344683e-61
## 1313  1.197422e-02
## 1343  2.353183e-49
## 1241  4.348629e-54
```

```
## 1223  8.802849e-03
## 1233  1.225492e-67
## 1143  1.912450e-70
## 2243  5.959151e-65
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
## [16,]    1    1    2    2
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
```

```
## [48,]    2    2    1    2
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

Bias-corrected estimates

Biased corrected estimates can be obtained by setting the option "BC = TRUE" (default "BC = FALSE"). The default is a bootstrap bias correction (default "method=`bootstrap' ") based on $10\,000$ bootstrap replicates (default "Bbias $= 10\,000$"). Alternatively, a jackknife bias correction can be obtained by setting the option "method=`jackknife"'. (The jachknife bias-correction ignores the optional argument "Bbias".)

The following code provides the bias-corrected MLEs for the dataset DATA based on 15 000 boostrap replicates:

```
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
##    lambda
## 0.9409539
##
## $haplotypes_frequencies
##               [,1]
## 1112  2.117201e-02
## 1121  5.859348e-02
## 1123  3.624001e-03
## 1131  1.146914e-02
## 1133  1.717813e-02
## 1221  2.423574e-02
## 1243  1.400871e-02
## 1311  7.862844e-03
## 1312  2.921807e-02
## 1321  1.377216e-01
## 1341  2.272151e-02
## 1323  3.358366e-02
## 1331  1.260695e-02
## 1333  2.323946e-02
## 1111 6.246350e-183
## 1122 5.691447e-151
## 1322 2.604422e-193
## 2121  3.538479e-02
## 2111  1.234348e-02
## 2133  3.692779e-02
## 2131  4.585024e-56
## 2123  1.442021e-89
```

```
## 2221  6.147224e-02
## 2222  1.711676e-02
## 2232  1.525365e-02
## 2231  1.547892e-02
## 2233  5.401264e-15
## 2311  2.558638e-02
## 2312  1.204467e-02
## 2313  2.188780e-02
## 2321  1.158500e-01
## 2333  5.073240e-02
## 2331  9.992021e-03
## 2323  1.255862e-02
## 2343  2.343080e-02
## 2141  2.608194e-02
## 2341  1.530016e-02
## 2143  1.021943e-02
## 2241  1.623833e-02
## 2211  6.102999e-08
## 2113  9.640364e-03
## 2213  1.044876e-56
## 2223  1.819387e-18
## 1113  6.894225e-03
## 1141  1.276602e-02
## 1222  1.002555e-02
## 1212  6.291714e-03
## 2212 1.630212e-202
## 1232 1.630212e-202
## 1231  7.502614e-61
## 1313  1.206916e-02
## 1343  4.478717e-49
## 1241  7.574051e-54
## 1223  1.147582e-02
## 1233  1.379400e-67
## 1143  4.577646e-70
## 2243  1.435973e-64
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
## [16,]    1    1    2    2
```

```
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
## [48,]    2    2    1    2
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

The bias-corrected MLEs with the 'jackknife' method and a plug-in value of the MOI parameter are obtained as follows:

```
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
## [1] 1
##
## $haplotypes_frequencies
##                 [,1]
```

```
## 1112   2.104026e-02
## 1121   5.836498e-02
## 1123   2.924205e-03
## 1131   7.952199e-03
## 1133   2.057565e-02
## 1221   2.637292e-02
## 1243   8.281213e-03
## 1311   7.865936e-03
## 1312   2.644752e-02
## 1321   1.462803e-01
## 1341   2.122421e-02
## 1323  -1.808922e-02
## 1331   2.551880e-02
## 1333  -1.834121e-02
## 1111  4.145459e-180
## 1122  8.030598e-149
## 1322  1.709275e-190
## 2121   3.325929e-02
## 2111   1.023583e-02
## 2133   3.742196e-02
## 2131   3.056572e-54
## 2123   1.045492e-87
## 2221   9.564064e-02
## 2222   1.576129e-02
## 2232   1.958688e-02
## 2231   1.404717e-02
## 2233   2.747309e-13
## 2311   4.403820e-02
## 2312   1.563780e-02
## 2313   1.358358e-02
## 2321   8.368002e-02
## 2333   1.399919e-01
## 2331  -2.653860e-03
## 2323   8.352781e-03
## 2343   2.710135e-02
## 2141   2.621678e-02
## 2341   1.417266e-02
## 2143   1.011664e-02
## 2241   1.279229e-02
## 2211   2.777061e-06
## 2113   1.246095e-02
## 2213   4.208844e-55
## 2223   9.599109e-17
## 1113   7.847298e-03
## 1141   1.017901e-02
## 1222   7.847644e-03
## 1212   1.173692e-02
## 2212  1.124954e-199
## 1232  1.124954e-199
## 1231   3.562640e-59
## 1313   2.136839e-02
## 1343   1.929610e-47
## 1241   3.565875e-52
## 1223   7.048240e-02
```

```
## 1233   1.004903e-65
## 1143   1.568209e-68
## 2243   4.886504e-63
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
## [16,]    1    1    2    2
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
## [48,]    2    2    1    2
```

```
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

Bootstrap confidence intervals

Moreover, equally-tailed $(1-\alpha)\times 100\%$ bootstrap-percentile confidence intervals (CIs) [**?**] are outputted alongside the estimates if the option "CI = TRUE" is specified (default "CI=FALSE"). The default are $B = 10\,000$ bootstrap repeats (default "B=10 000") and $\alpha = 0.05$ (default "alpha=0.05"). In this case, the MOI parameter estimate is a vector that contains the MLE, and the upper and lower confidence points, except a plug-in estimate is provided. The haplotype frequencies are provided as an array with 3 columns, where the first column provides the estimates, and the second and third the upper and lower confidence points, respectively.

To obtain the estimates of the MOI parameter and haplotype frequencies with their corresponding 95% confidence intervals based on 15 000 bootstrap replicates, one should run the following code:

```
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
##                2.5%     97.5%
## 0.9030599 0.8147888 1.1195641
##
## $haplotypes_frequencies
##                              2.5%        97.5%
## 1112  1.621991e-02  3.017636e-03 4.089470e-02
## 1121  5.662133e-02  2.550345e-02 1.004115e-01
## 1123  1.004166e-02  0.000000e+00 4.518934e-02
## 1131  1.290682e-02  7.438009e-24 4.395086e-02
## 1133  1.414079e-02  1.498891e-09 3.194084e-02
## 1221  2.502605e-02  1.022560e-02 6.841769e-02
## 1243  8.766197e-03  0.000000e+00 2.822228e-02
## 1311  9.622878e-03  1.763213e-92 1.502974e-02
## 1312  2.886627e-02  7.820172e-03 5.043365e-02
## 1321  1.442050e-01  7.912979e-02 1.438373e-01
## 1341  2.224698e-02  2.043287e-03 4.676423e-02
## 1323  2.174462e-02  2.035195e-03 6.493588e-02
## 1331  1.605193e-02  3.153786e-64 2.346484e-02
## 1333  1.632423e-02  1.281883e-02 8.715079e-02
## 1111 6.246350e-183  0.000000e+00 2.280239e-72
## 1122 5.691447e-151  0.000000e+00 2.280239e-72
## 1322 2.604422e-193  0.000000e+00 1.484005e-73
## 2121  3.812513e-02  1.053444e-02 7.839116e-02
## 2111  6.244674e-03  0.000000e+00 2.444078e-02
## 2133  3.465128e-02  9.874639e-53 7.421984e-02
## 2131  4.585024e-56  4.399370e-80 1.877681e-02
```

```
## 2123  1.442021e-89  0.000000e+00 4.060172e-31
## 2221  6.241974e-02  1.061338e-02 1.086927e-01
## 2222  1.667291e-02  0.000000e+00 2.253940e-02
## 2232  1.344530e-02  1.781156e-03 4.422355e-02
## 2231  1.300739e-02  9.216091e-54 3.588970e-02
## 2233  5.401264e-15 1.326123e-112 3.442752e-02
## 2311  2.884877e-02  1.953346e-54 7.374667e-02
## 2312  1.435814e-02  3.309262e-03 3.104265e-02
## 2313  1.923153e-02  9.380481e-03 3.694464e-02
## 2321  1.181361e-01  6.942905e-02 1.731926e-01
## 2333  6.419634e-02  2.925446e-33 9.814590e-02
## 2331  7.108918e-03  1.456265e-34 2.572674e-02
## 2323  5.614757e-03  1.782575e-79 2.954590e-02
## 2343  2.832524e-02  5.263614e-10 2.547182e-02
## 2141  2.517656e-02  4.455277e-50 4.036218e-02
## 2341  1.883940e-02  3.350963e-48 4.024499e-02
## 2143  8.154858e-03  0.000000e+00 2.062635e-02
## 2241  2.159266e-02  6.322824e-26 2.827766e-02
## 2211  6.102999e-08  0.000000e+00 1.431560e-02
## 2113  1.124325e-02  7.299470e-27 2.919215e-02
## 2213  1.044876e-56  0.000000e+00 6.663777e-08
## 2223  1.819387e-18  1.334681e-79 1.390215e-02
## 1113  9.091810e-03  0.000000e+00 2.919215e-02
## 1141  1.355581e-02 1.963258e-304 2.952343e-02
## 1222  8.471195e-03  0.000000e+00 1.415688e-02
## 1212  5.420756e-03  0.000000e+00 1.457896e-02
## 2212 1.630212e-202  0.000000e+00 8.366406e-03
## 1232 1.630212e-202  0.000000e+00 8.366406e-03
## 1231  7.502614e-61  0.000000e+00 3.636187e-03
## 1313  1.323361e-02  3.874104e-81 2.790564e-02
## 1343  4.478717e-49 1.815479e-293 5.031388e-03
## 1241  7.574051e-54 2.035012e-217 6.505116e-03
## 1223  1.484892e-02 9.013752e-221 1.509500e-02
## 1233  1.379400e-67  0.000000e+00 4.666472e-22
## 1143  4.577646e-70  0.000000e+00 2.140550e-30
## 2243  1.435973e-64  0.000000e+00 1.143886e-10
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
```

```
## [16,]    1    1    2    2
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
## [48,]    2    2    1    2
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

The following code provides the estimates with 90% CIs based on 20 000 bootstrap repeats:

```
calculateMaximumLikelihoodEstimatesWithAddOns(datasetStandard[[1]][,markers], datasetStandard[[3]][mark
```

```
## $lambda
##                  5%       95%
## 0.9226857 0.8678825 1.1592490
##
## $haplotypes_frequencies
```

```
##                              5%          95%
## 1112  2.501760e-02  6.457533e-03 6.051031e-02
## 1121  6.659734e-02  3.405667e-02 1.057151e-01
## 1123  8.644535e-03 2.122262e-122 2.375365e-02
## 1131  1.204386e-02  0.000000e+00 2.596665e-02
## 1133  1.145501e-02  0.000000e+00 3.044447e-02
## 1221  2.507386e-02  7.739312e-03 5.089666e-02
## 1243  1.234275e-02  0.000000e+00 2.695205e-02
## 1311  6.169786e-03  0.000000e+00 1.599113e-02
## 1312  3.545312e-02  1.062296e-02 4.846415e-02
## 1321  1.461108e-01  9.834997e-02 1.730760e-01
## 1341  1.771481e-02  1.044438e-02 5.884071e-02
## 1323  2.045031e-02  7.779898e-03 5.512222e-02
## 1331  1.900005e-02 3.385337e-151 2.595869e-02
## 1333  1.113583e-02  8.758015e-40 6.262769e-02
## 1111 6.246350e-183  0.000000e+00 1.761004e-35
## 1122 5.691447e-151  0.000000e+00 1.768666e-28
## 1322 2.604422e-193  0.000000e+00 4.428472e-33
## 2121  3.456065e-02  8.814374e-03 6.767834e-02
## 2111  8.969445e-03  0.000000e+00 2.326812e-02
## 2133  3.523797e-02  1.826996e-20 7.310434e-02
## 2131  4.585024e-56 8.848097e-131 1.553041e-02
## 2123  1.442021e-89 1.413170e-251 1.319133e-20
## 2221  6.081908e-02  9.168206e-03 7.648551e-02
## 2222  1.374605e-02  0.000000e+00 3.491303e-02
## 2232  1.743076e-02  3.462807e-03 3.209740e-02
## 2231  1.952319e-02 5.220300e-250 4.872410e-02
## 2233  5.401264e-15 2.732703e-105 2.296789e-02
## 2311  2.922122e-02  1.091778e-65 6.098623e-02
## 2312  1.178710e-02  0.000000e+00 3.230494e-02
## 2313  2.559351e-02  0.000000e+00 4.683238e-02
## 2321  1.075994e-01  6.690167e-02 1.539179e-01
## 2333  6.308539e-02  1.593743e-53 8.736437e-02
## 2331  4.471493e-03  9.726506e-87 2.917117e-02
## 2323  9.076581e-03  1.722233e-71 3.912914e-02
## 2343  2.409350e-02  1.966712e-03 5.174652e-02
## 2141  2.206349e-02  2.881861e-31 3.412446e-02
## 2341  1.168954e-02  2.014176e-31 5.234434e-02
## 2143  8.746034e-03  2.158611e-76 2.939949e-02
## 2241  1.352120e-02  7.012950e-38 4.108869e-02
## 2211  6.102999e-08  3.707822e-17 2.367115e-02
## 2113  1.252606e-02  0.000000e+00 3.204260e-02
## 2213  1.044876e-56  0.000000e+00 1.016276e-09
## 2223  1.819387e-18 2.133709e-178 9.860227e-03
## 1113  9.555259e-03  0.000000e+00 2.452674e-02
## 1141  1.103062e-02  1.687677e-44 3.079318e-02
## 1222  6.867750e-03  0.000000e+00 2.433775e-02
## 1212  8.940157e-03  0.000000e+00 1.417219e-02
## 2212 1.630212e-202  0.000000e+00 4.184328e-03
## 1232 1.630212e-202  0.000000e+00 4.184328e-03
## 1231  7.502614e-61  0.000000e+00 7.236929e-19
## 1313  1.595226e-02  0.000000e+00 2.875118e-02
## 1343  4.478717e-49 1.314415e-192 1.546184e-13
## 1241  7.574051e-54  0.000000e+00 2.904249e-09
```

```
## 1223  1.400732e-02  0.000000e+00 9.820595e-03
## 1233  1.379400e-67  0.000000e+00 3.928022e-28
## 1143  4.577646e-70  0.000000e+00 1.679426e-14
## 2243  1.435973e-64  0.000000e+00 7.305524e-05
##
## $detected_haplotypes
##       [,1] [,2] [,3] [,4]
##  [1,]    1    1    1    2
##  [2,]    1    1    2    1
##  [3,]    1    1    2    3
##  [4,]    1    1    3    1
##  [5,]    1    1    3    3
##  [6,]    1    2    2    1
##  [7,]    1    2    4    3
##  [8,]    1    3    1    1
##  [9,]    1    3    1    2
## [10,]    1    3    2    1
## [11,]    1    3    4    1
## [12,]    1    3    2    3
## [13,]    1    3    3    1
## [14,]    1    3    3    3
## [15,]    1    1    1    1
## [16,]    1    1    2    2
## [17,]    1    3    2    2
## [18,]    2    1    2    1
## [19,]    2    1    1    1
## [20,]    2    1    3    3
## [21,]    2    1    3    1
## [22,]    2    1    2    3
## [23,]    2    2    2    1
## [24,]    2    2    2    2
## [25,]    2    2    3    2
## [26,]    2    2    3    1
## [27,]    2    2    3    3
## [28,]    2    3    1    1
## [29,]    2    3    1    2
## [30,]    2    3    1    3
## [31,]    2    3    2    1
## [32,]    2    3    3    3
## [33,]    2    3    3    1
## [34,]    2    3    2    3
## [35,]    2    3    4    3
## [36,]    2    1    4    1
## [37,]    2    3    4    1
## [38,]    2    1    4    3
## [39,]    2    2    4    1
## [40,]    2    2    1    1
## [41,]    2    1    1    3
## [42,]    2    2    1    3
## [43,]    2    2    2    3
## [44,]    1    1    1    3
## [45,]    1    1    4    1
## [46,]    1    2    2    2
## [47,]    1    2    1    2
```

```
## [48,]    2    2    1    2
## [49,]    1    2    3    2
## [50,]    1    2    3    1
## [51,]    1    3    1    3
## [52,]    1    3    4    3
## [53,]    1    2    4    1
## [54,]    1    2    2    3
## [55,]    1    2    3    3
## [56,]    1    1    4    3
## [57,]    2    2    4    3
##
## $used_sample_size
## [1] 82
```

Linkage disequilibrium estimates

The function ld(<DATA>,...) derives LD measures from the output of the function mle(<DATA>,...). The function outputs the four LD measures D', r^2, Q^*, and the ALD measures $W_{A|B}$ and $W_{B|A}$. Moreover, the option to output the $(1-\alpha)$-level bootstrap confidence intervals for the LD estimates or bias-corrected estimates are available and are used as with the function mle(<DATA>,...). Estimation of LD with a 95% confidence interval is done in the following code snippet:

```
markersPair <- c(4,4)
calculatePairwiseLDWithAddons(datasetStandard,markersPair, idExists = FALSE)
```

```
##   D'  r^2
## 0.10 0.01
```

```
calculatePairwiseLDWithAddons(datasetStandard,markersPair, idExists = FALSE, isConfidenceInterval=TRUE,r
```

```
##          2.5% 97.5%
## D'  0.10 0.04 0.261
## r^2 0.01 0.00 0.040
```