

LD_SNPS_RESISTIN

KIRABO GLORIA

2023-05-17

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr  1.0.1
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: package 'stringr' was built under R version 4.2.3
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.2.3
```

```
library(haplo.stats)
```

```
## Warning: package 'haplo.stats' was built under R version 4.2.3
```

```
## Loading required package: arsenal
```

```
## Warning: package 'arsenal' was built under R version 4.2.3
```

```
library(dplyr)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:arsenal':
##
##   set_attr
##
## The following object is masked from 'package:purrr':
##
```

```

##      set_names
##
## The following object is masked from 'package:tidyr':
##
##      extract

library(genetics)

## Warning: package 'genetics' was built under R version 4.2.3

## Loading required package: combinat
##
## Attaching package: 'combinat'
##
## The following object is masked from 'package:utils':
##
##      combn
##
## Loading required package: gdata

## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2

## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2

## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
##
## gdata: Unable to load perl libraries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
##
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libraries needed to support Excel XLS and XLSX formats.
##
## Attaching package: 'gdata'
##
## The following objects are masked from 'package:dplyr':
##
##      combine, first, last
##
## The following object is masked from 'package:purrr':
##
##      keep
##
## The following object is masked from 'package:stats':
##
##      nobs
##
## The following object is masked from 'package:utils':

```

```
##
##      object.size
##
## The following object is masked from 'package:base':
##
##      startsWith
##
## Loading required package: gtools

## Warning: package 'gtools' was built under R version 4.2.3

## Loading required package: MASS

## Warning: package 'MASS' was built under R version 4.2.3

##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##      select
##
## Loading required package: mvtnorm
##
##
## NOTE: THIS PACKAGE IS NOW OBSOLETE.
##
##
##
## The R-Genetics project has developed an set of enhanced genetics
## packages to replace 'genetics'. Please visit the project homepage
## at http://rgenetics.org for informtion.
##
##
##
## Attaching package: 'genetics'
##
## The following object is masked from 'package:haplo.stats':
##
##      locus
##
## The following objects are masked from 'package:base':
##
##      %in%, as.factor, order
```

```
##STEP ONE-READING IN THE DATASET FROM THE LINK
```

```
fams <- read.delim("http://www.biostat.umn.edu/~cavanr/FMS_data.txt",header = T,
  sep = "\t")
```

```
#Vieing the first five rows and 10columns of the dataset
```

```
fams[1:5,1:10]
```

```
##      id acdc_rs1501299 ace_id actn3_r577x actn3_rs540874 actn3_rs1815739
## 1 FA-1801          CA      DD          CC          GG          CC
## 2 FA-1802          CA      ID          CT          GA          TC
## 3 FA-1803          CA      ID          CT          GA          TC
## 4 FA-1804          CC      DD          CT          GA          TC
## 5 FA-1805          CA      ID          CC          GG          CC
##      actn3_1671064 ardb1_1801253 adrb2_1042713 adrb2_1042714
## 1          AA          <NA>          GA          CG
## 2          GA          <NA>          GA          CC
## 3          GA          <NA>          GA          CG
## 4          GA          <NA>          AA          CC
## 5          AA          <NA>          GA          CG
```

##SELECTING OUT COLUMNS WITH RESISTIN GENES

```
fams_restn <- fams[grepl("^resistin", names(fams))]
head(fams_restn)
```

```
##      resistin_c30t resistin_c398t resistin_g540a resistin_c980g resistin_c180g
## 1          CC          CC          GG          GG          CC
## 2          CC          TT          AA          CG          GG
## 3          CC          CC          GG          CG          CC
## 4          CC          CC          GA          CG          CG
## 5          CT          CC          GG          CC          CC
## 6          CC          CC          GG          CG          CC
##      resistin_a537c
## 1          AA
## 2          AA
## 3          AA
## 4          AA
## 5          AA
## 6          AA
```

##Counting the number of generated columns

```
restn_col <- ncol(fams_restn)
cat("The columns with resistin gene are:", restn_col)
```

```
## The columns with resistin gene are: 6
```

##IDENTIFYING UNIQUE SNPS IN THE COLUMNS

##STEP ONE CREATE A VECTOR

```
restn_snp <- unlist(fams_restn)
head(restn_snp)
```

```
## resistin_c30t1 resistin_c30t2 resistin_c30t3 resistin_c30t4 resistin_c30t5
##          "CC"          "CC"          "CC"          "CC"          "CT"
## resistin_c30t6
##          "CC"
```

##STEP TWO:IDENTIFYING UNIQUE VALUES IN THE VVECTOR

```
num_snps_restn <- length(unique(restn_snp))  
cat("The total number of snps in the restin genes is :",num_snps_restn)
```

```
## The total number of snps in the restin genes is : 9
```

##CREATING A GENOTYPE OBJECT FOR OUR GENES OF THE RESISTIN GENE

```
geno_restn <- as.data.frame(lapply(fams_restn, genotype, sep=""))  
geno_restn[1:5,]
```

```
##   resistin_c30t resistin_c398t resistin_g540a resistin_c980g resistin_c180g  
## 1           C/C           C/C           G/G           G/G           C/C  
## 2           C/C           T/T           A/A           C/G           G/G  
## 3           C/C           C/C           G/G           C/G           C/C  
## 4           C/C           C/C           G/A           C/G           C/G  
## 5           C/T           C/C           G/G           C/C           C/C  
##   resistin_a537c  
## 1           A/A  
## 2           A/A  
## 3           A/A  
## 4           A/A  
## 5           A/A
```

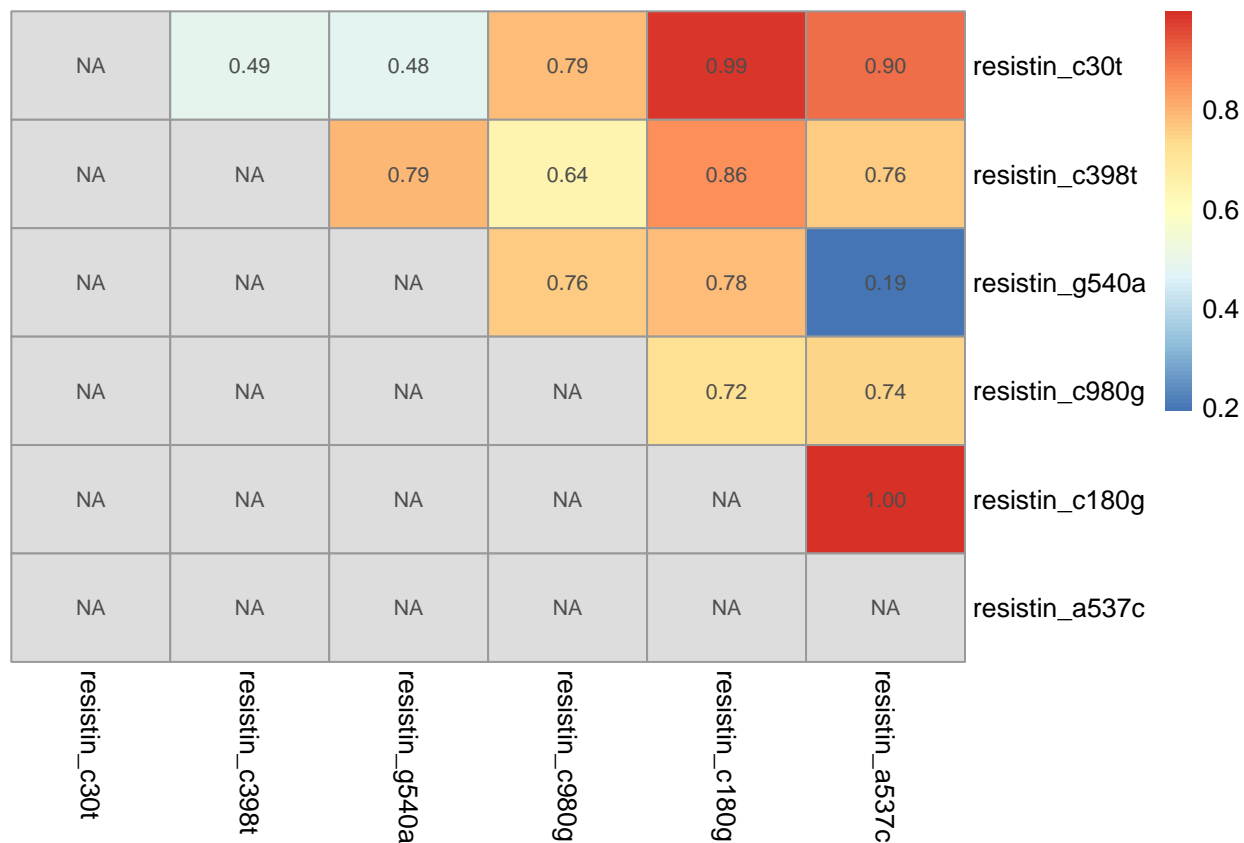
##Calculating D' OF THE SNPS IN THE RESISTIN GENE

```
restn_D <- LD(geno_restn)$`D`  
restn_D
```

```
##           resistin_c30t resistin_c398t resistin_g540a resistin_c980g  
## resistin_c30t           NA      0.4855449      0.4810285      0.7877728  
## resistin_c398t           NA           NA      0.7944463      0.6403467  
## resistin_g540a           NA           NA           NA      0.7586427  
## resistin_c980g           NA           NA           NA           NA  
## resistin_c180g           NA           NA           NA           NA  
## resistin_a537c           NA           NA           NA           NA  
##           resistin_c180g resistin_a537c  
## resistin_c30t      0.9850864      0.9021448  
## resistin_c398t      0.8581316      0.7633220  
## resistin_g540a      0.7833792      0.1929656  
## resistin_c980g      0.7183456      0.7441364  
## resistin_c180g           NA      0.9983006  
## resistin_a537c           NA           NA
```

##GENERATING A HEATMAP TO SHOW THE D' OF SNPS

```
pheatmap(restn_D, cluster_cols = FALSE, cluster_rows = FALSE, display_numbers = TRUE)
```



##From the heatmap, it shows that the SNP at resistin_c180g is in high LD with resistin_a537c
 ## Also resistin_c30t is in high LD with resistin_c180g.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.