

Antonin Rosa

Hendrik Wellmann

Practical 3 Statistical Genetics

Linkage disequilibrium

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

Question 1

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

```
library(genetics)
```

```
## Loading required package: combinat
```

```
##
```

```
## Attaching package: 'combinat'
```

```
## The following object is masked from 'package:utils':
```

```
##
```

```
##      combn
```

```
## Loading required package: gdata
```

```
##
```

```
## Attaching package: 'gdata'
```

```
## 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
```

```
## Loading required package: MASS
```

```
## 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
library(HardyWeinberg)

## Loading required package: mice
##
## Attaching package: 'mice'
## The following object is masked from 'package:stats':
##
##      filter
## The following objects are masked from 'package:base':
##
##      cbind, rbind
## Loading required package: Rsolnp
## Loading required package: nnet
data <- read.table("FOXP2/FOXP2.dat", header = TRUE, sep = " ")
num_individuals <- nrow(data) - 1
num_snps <- ncol(data) - 1

cat("Number of individuals:", num_individuals, "\n")

## Number of individuals: 103
cat("Number of SNPs:", num_snps, "\n")

## Number of SNPs: 543
missing_percentage <- mean(is.na(data)) * 100
cat("Percentage of missing data:", missing_percentage, "%\n")

## Percentage of missing data: 0 %

g1 <- genotype(data$rs34684677)
g2 <- genotype(data$rs2894715)
LD(g1,g2)

```

Question 2

```

##
## Pairwise LD
## -----
##              D          D'          Corr
## Estimates: -0.05493703 0.9986536 -0.3144048
##
##              X^2      P-value      N
## LD Test: 20.56088 5.77645e-06 104

```

The SNPs rs34684677 and rs2894715 show a strong genetic linkage, meaning that variations in one SNP are highly likely to be associated with variations in the other. The high D' value (0.9986536) indicates a strong linkage. The moderate negative correlation (Corr = -0.3144048) suggests that variations in allele frequencies tend to occur in opposite directions. In summary, there's a significant and non-random association between alleles of rs34684677 and rs2894715, with a tendency for opposite variations.

Question 3 Low LD ($D = -0.05493703$) suggests weak association between SNPs rs34684677 and rs2894715. Haplotypes are likely formed independently. Estimate haplotype frequencies from genotype frequencies. The most common haplotype has the highest estimated frequency. Provide genotype frequencies for detailed calculation

```
snp1 <- data$rs34684677
snp2 <- data$rs2894715
genotype_matrix <- cbind(substr(snp1, 1, 1), substr(snp1, 3, 3),
                          substr(snp2, 1, 1), substr(snp2, 3, 3))

snpnames <- c("snp1", "snp2")
haplo_freq <- haplo.em(genotype_matrix, locus.label=snpnames <- c("snp1", "snp2"))
print(haplo_freq)
```

```
## =====
##                                     Haplotypes
## =====
##   snp1 snp2 hap.freq
## 1    G    G  0.33654
## 2    G    T  0.50000
## 3    T    G  0.00000
## 4    T    T  0.16346
## =====
##                                     Details
## =====
## lnlike =  -164.8458
## lr stat for no LD =  18.69923 , df =  0 , p-val =  NA
```

The answer is SNP1=G and SNP2=T.

```
bim_file <- "FOXP2/FOXP2.bim"
bim_data <- read.table(bim_file, header = FALSE, stringsAsFactors = FALSE)

genotype_matrix <- data[,-1]
for (col in colnames(genotype_matrix)) {
  genotype_matrix[[col]] <- genotype(genotype_matrix[[col]])
}

calculate_MAF <- function(genotype_counts) {
  allele1 <- 2 * genotype_counts[[1]] + genotype_counts[[2]]
  allele2 <- 2 * genotype_counts[[3]] + genotype_counts[[2]]
  return (min(allele1, allele2) / (allele2 + allele1))
}

filter_index_MAF <- c()
res <- 0
for (i in 1:ncol(genotype_matrix)) {
  alleles <- c(paste(bim_data$V5[[i]], "/", bim_data$V6[[i]], sep = ""))
  count_genotype <- MakeCounts(genotype_matrix[[i]], alleles = alleles, sep="/")
}
```

```

MAF <- calculate_MAF(count_gentoype)
if (MAF < 0.35) {
  filter_index_MAF <- c(filter_index_MAF, i)
}
p_value <- HWE.chisq(genotype_matrix[[i]])
p_value <- p_value$p.value
if (p_value < 0.05) {
  res <- res + 1
}
}
print(res)

```

Question 4

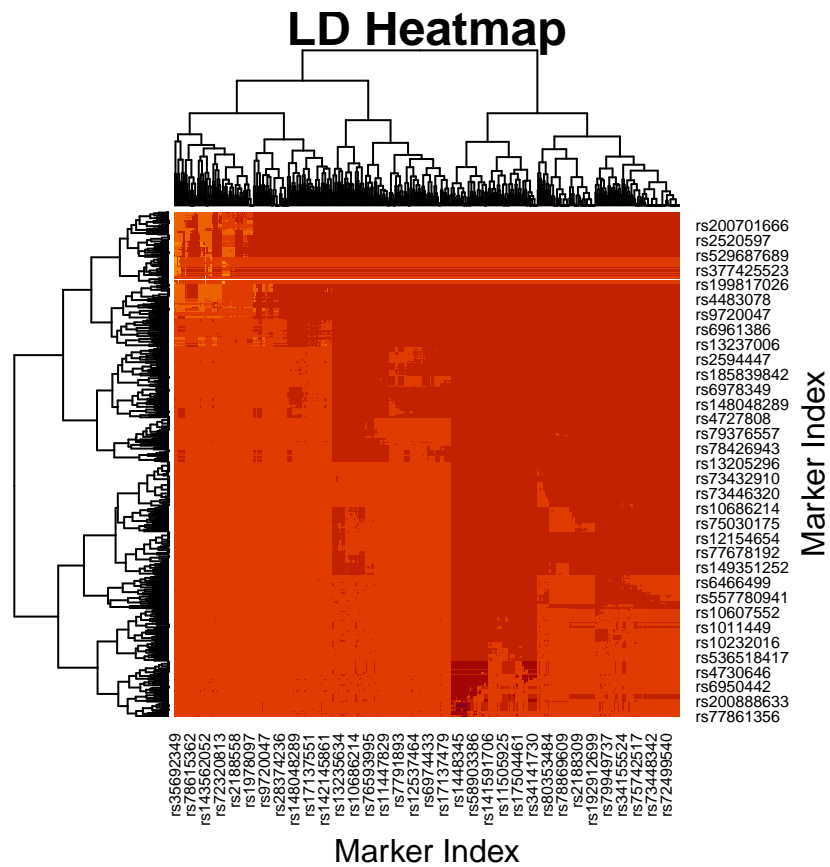
```
## [1] 18
```

We can reject 18 variants.

```

ld_matrix <- LD(genotype_matrix)
r2_matrix <- ld_matrix$`R^2`
r2_matrix[is.na(r2_matrix)] <- 1
heatmap(r2_matrix, main = "LD Heatmap", xlab = "Marker Index", ylab = "Marker Index")

```



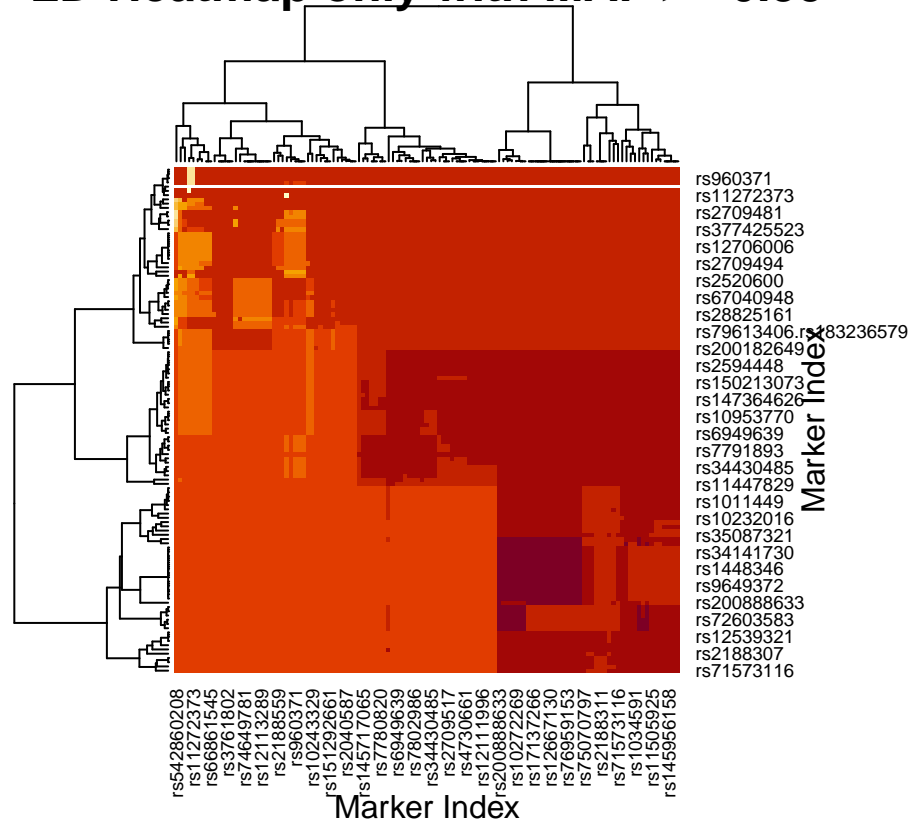
Question 5

```

r2_matrix <- r2_matrix[!filter_index_MAF, !filter_index_MAF]
heatmap(r2_matrix, main = "LD Heatmap only with MAF >= 0.35", xlab = "Marker Index", ylab = "Marker Index")

```

LD Heatmap only with MAF ≥ 0.35



Question 6

Question 7

Haplotype estimation

```
data <- read.table("APOE/APOE.dat", header = TRUE, sep = " ")
num_individuals <- nrow(data) - 1
num_snps <- ncol(data) - 1
cat("Number of individuals:", num_individuals, "\n")
```

Question 1

```
## Number of individuals: 106
```

```
cat("Number of SNPs:", num_snps, "\n")
```

```
## Number of SNPs: 162
```

```
missing_percentage <- mean(is.na(data)) * 100
```

```
cat("Percentage of missing data:", missing_percentage, "%\n")
```

```
## Percentage of missing data: 0 %
```

Question 2 We can theoretically find 2^{544} .

```

n_line <- nrow(data)
n_col <- 2 * ncol(data)
Geno <- matrix(0, nrow = n_line, ncol = n_col)

for (i in 1:ncol(data)) {
  col_start <- (i - 1) * 2 + 1
  Geno[, col_start] <- substr(data[, i], 1, 1)
  Geno[, col_start + 1] <- substr(data[, i], 3, 3)
}

options(max.print = 100000)

haplo_freq <- haplo.em(Geno)
print(haplo_freq)

```

Question 3

```

## =====
##                                     Haplotypes
## =====
##   loc-1 loc-2 loc-3 loc-4 loc-5 loc-6 loc-7 loc-8 loc-9 loc-10 loc-11 loc-12
## 1      2      C      C      C      G      C      A      C      G      G      C      T
## 2      2      C      C      C      G      C      A      C      G      G      C      T
## 3      2      C      C      C      G      C      A      C      G      G      C      T
## 4      2      C      C      C      G      C      A      C      G      G      C      T
## 5      2      C      C      C      G      C      A      C      G      G      C      T
## 6      2      C      C      C      G      C      A      C      G      G      C      T
## 7      2      C      C      C      G      C      G      C      A      C      C      T
## 8      2      C      C      C      G      C      G      C      G      C      C      T
## 9      2      C      C      C      G      C      G      C      G      C      C      T
## 10     2      C      C      C      G      C      G      C      G      C      C      T
## 11     2      C      C      C      G      C      G      C      G      C      C      T
## 12     2      C      C      C      G      C      G      C      G      C      C      T
## 13     2      C      C      C      G      C      G      C      G      C      C      T
## 14     2      C      C      C      G      C      G      C      G      C      C      T
## 15     2      C      C      C      G      C      G      C      G      C      C      T
## 16     2      C      C      C      G      C      G      C      G      C      C      T
## 17     2      C      C      C      G      C      G      C      G      C      C      T
## 18     2      C      C      C      G      C      G      C      G      G      C      T
## 19     2      C      C      C      G      C      G      C      G      G      C      T
## 20     2      C      C      C      G      C      G      C      G      G      C      T
## 21     2      C      C      C      G      C      G      C      G      G      C      T
## 22     2      C      C      C      G      C      G      C      G      G      C      T
## 23     2      C      C      C      G      C      G      C      G      G      C      T
## 24     2      C      C      C      G      C      G      C      G      G      C      T
## 25     2      C      C      C      G      C      G      C      G      G      C      T
## 26     2      C      C      C      G      C      G      C      G      G      C      T
## 27     2      C      C      C      G      C      G      C      G      G      C      T
## 28     2      C      C      C      G      C      G      C      G      G      C      T
## 29     2      C      C      C      G      C      G      C      G      G      C      T
## 30     2      C      C      C      G      C      G      C      G      G      C      T
## 31     2      C      C      C      G      C      G      C      G      G      C      T
## 32     2      C      C      C      G      C      G      C      G      G      C      T
## 33     2      C      C      C      G      C      G      C      G      G      C      T

```

## 34	2	C	C	C	G	C	G	C	G	G	C	T
## 35	2	G	C	C	G	C	G	C	G	G	C	T
## 36	N	C	C	C	G	C	A	C	G	G	C	T
## 37	N	C	C	C	G	C	G	C	G	C	C	T
## 38	N	C	C	C	G	C	G	C	G	G	C	T
## 39	N	C	C	C	G	C	G	C	G	G	C	T
## 40	N	C	C	C	G	C	G	C	G	G	C	T
## 41	N	C	C	C	G	C	G	C	G	G	C	T
##	loc-13	loc-14	loc-15	loc-16	loc-17	loc-18	loc-19	loc-20	loc-21	loc-22	loc-23	
## 1	G	C	C	C	G	G	G	C	T	C	A	
## 2	G	C	C	C	G	G	G	C	T	C	A	
## 3	G	C	C	C	G	G	G	C	T	C	A	
## 4	G	C	C	C	G	G	G	C	T	C	A	
## 5	G	C	C	C	G	G	G	C	T	C	A	
## 6	G	C	C	C	G	G	G	C	T	C	A	
## 7	G	C	C	C	G	G	G	C	T	C	A	
## 8	G	A	C	C	G	G	G	C	T	C	A	
## 9	G	C	C	C	A	G	G	C	T	C	A	
## 10	G	C	C	C	A	G	G	C	T	C	A	
## 11	G	C	C	C	G	G	A	C	T	C	A	
## 12	G	C	C	C	G	G	A	C	T	C	A	
## 13	G	C	C	C	G	G	A	C	T	C	A	
## 14	G	C	C	C	G	G	A	C	T	C	A	
## 15	G	C	C	C	G	G	G	C	T	C	A	
## 16	G	C	C	C	G	G	G	C	T	C	A	
## 17	G	C	C	C	G	G	G	C	T	C	A	
## 18	G	C	C	C	G	G	G	C	T	C	A	
## 19	G	C	C	C	G	G	G	C	T	C	A	
## 20	G	C	C	C	G	G	G	C	T	C	A	
## 21	G	C	C	C	G	G	G	C	T	C	A	
## 22	G	C	C	C	G	G	G	C	T	C	A	
## 23	G	C	C	C	G	G	G	C	T	C	A	
## 24	G	C	C	C	G	G	G	C	T	C	A	
## 25	G	C	C	C	G	G	G	C	T	C	A	
## 26	G	C	C	C	G	G	G	C	T	C	A	
## 27	G	C	C	C	G	G	G	C	T	C	A	
## 28	G	C	C	C	G	G	G	C	T	C	A	
## 29	G	C	C	C	G	G	G	C	T	C	A	
## 30	G	C	C	C	G	G	G	C	T	C	A	
## 31	G	C	C	C	G	G	G	C	T	C	A	
## 32	G	C	C	C	G	G	G	C	T	C	A	
## 33	G	C	C	C	G	G	G	C	T	C	A	
## 34	G	C	C	C	G	G	G	C	T	C	A	
## 35	G	C	C	C	G	G	G	C	T	C	A	
## 36	G	C	C	C	G	G	G	C	T	C	A	
## 37	G	C	C	C	A	G	G	C	T	C	A	
## 38	G	C	C	C	G	G	G	C	T	C	A	
## 39	G	C	C	C	G	G	G	C	T	C	A	
## 40	G	C	C	C	G	G	G	C	T	C	A	
## 41	G	C	C	C	G	G	G	C	T	C	A	
##	loc-24	loc-25	loc-26	loc-27	loc-28	loc-29	loc-30	loc-31	loc-32	loc-33	loc-34	
## 1	T	T	C	T	T	A	A	T	C	A	C	
## 2	T	T	C	T	T	C	A	T	C	A	C	
## 3	T	T	C	T	T	C	A	T	C	A	C	

## 4	T	T	C	T	T	C	A	T	C	A	C
## 5	T	T	C	T	T	C	A	T	C	A	C
## 6	T	T	C	T	T	C	A	T	C	A	C
## 7	T	T	C	T	C	C	A	C	C	A	C
## 8	T	T	C	T	C	C	A	C	C	A	C
## 9	T	T	C	T	C	C	A	C	C	A	C
## 10	T	T	C	T	C	C	A	C	C	A	C
## 11	T	T	C	T	C	C	A	C	C	A	C
## 12	T	T	C	T	C	C	A	C	C	A	C
## 13	T	T	C	T	C	C	A	C	C	A	C
## 14	T	T	C	T	C	C	A	C	C	A	C
## 15	T	T	C	T	C	C	A	C	C	A	C
## 16	T	T	C	T	C	C	A	C	C	A	C
## 17	T	T	C	T	C	C	A	C	C	A	C
## 18	T	T	C	T	T	C	A	T	C	A	C
## 19	T	T	C	T	T	C	A	T	C	A	C
## 20	T	T	C	T	T	C	A	T	C	A	C
## 21	T	T	C	T	T	C	A	T	C	A	C
## 22	T	T	C	T	T	C	A	T	C	A	C
## 23	T	T	C	T	T	C	A	T	C	A	C
## 24	T	T	C	T	T	C	A	T	C	A	C
## 25	T	T	C	T	T	C	A	T	C	A	C
## 26	T	T	C	T	T	C	A	T	C	A	C
## 27	T	T	C	T	T	C	A	T	C	A	C
## 28	T	T	C	T	T	C	A	T	C	A	C
## 29	T	T	C	T	T	C	A	T	C	A	C
## 30	T	T	C	T	T	C	A	T	C	A	C
## 31	T	T	C	T	T	C	A	T	C	A	C
## 32	T	T	C	T	T	C	A	T	C	A	C
## 33	T	T	C	T	T	C	A	T	C	A	C
## 34	T	T	C	T	T	C	A	T	C	A	C
## 35	T	T	C	T	T	C	A	T	C	A	C
## 36	T	T	C	T	T	C	A	T	C	A	C
## 37	T	T	C	T	C	C	A	C	C	A	C
## 38	T	T	C	T	T	C	A	T	C	A	C
## 39	T	T	C	T	T	C	A	T	C	A	C
## 40	T	T	C	T	T	C	A	T	C	A	C
## 41	T	T	C	T	T	C	A	T	C	A	C
##	loc-35	loc-36	loc-37	loc-38	loc-39	loc-40	loc-41	loc-42	loc-43	loc-44	loc-45
## 1	C	T	T	A	G	C	C	C	A	C	C
## 2	C	T	C	A	G	C	C	C	A	C	C
## 3	C	T	C	A	G	C	C	C	A	C	C
## 4	C	T	C	A	G	C	C	C	A	C	C
## 5	C	T	C	A	G	C	C	C	A	C	C
## 6	C	T	C	A	G	C	C	C	A	C	C
## 7	C	T	C	G	G	T	C	C	C	C	C
## 8	C	T	C	G	G	T	C	C	C	C	C
## 9	C	T	C	G	G	T	C	C	C	C	C
## 10	C	T	C	G	G	T	C	C	C	C	C
## 11	C	T	C	G	G	T	C	C	C	C	C
## 12	C	T	C	G	G	T	C	C	C	C	C
## 13	C	T	C	G	G	T	C	C	C	C	C
## 14	C	T	C	G	G	T	C	C	C	C	C
## 15	C	T	C	G	G	T	C	C	C	C	C

## 16	C	T	C	G	G	T	C	C	C	C	C
## 17	C	T	T	G	G	T	C	C	C	C	C
## 18	C	T	C	A	G	C	C	C	A	C	C
## 19	C	T	C	A	G	C	C	C	A	C	C
## 20	C	T	C	A	G	C	C	C	A	C	C
## 21	C	T	C	A	G	C	C	C	A	C	C
## 22	C	T	C	A	G	C	C	C	A	C	C
## 23	C	T	C	A	G	C	C	C	A	C	C
## 24	C	T	C	A	G	C	C	C	A	C	C
## 25	C	T	C	A	G	C	C	C	A	C	C
## 26	C	T	C	A	G	C	C	C	A	C	C
## 27	C	T	C	A	G	C	C	C	A	C	C
## 28	C	T	C	A	G	C	C	C	A	C	C
## 29	C	T	C	A	G	C	C	C	A	C	C
## 30	C	T	C	A	G	C	C	C	A	C	C
## 31	C	T	C	A	G	C	C	C	A	C	C
## 32	C	T	T	A	G	C	C	C	A	C	C
## 33	C	T	T	A	G	C	C	C	A	C	C
## 34	C	T	T	A	G	C	C	C	A	C	C
## 35	C	T	C	A	G	C	C	C	A	C	C
## 36	C	T	C	A	G	C	C	C	A	C	C
## 37	C	T	C	G	G	T	C	C	C	C	C
## 38	C	T	C	A	G	C	C	C	A	C	C
## 39	C	T	C	A	G	C	C	C	A	C	C
## 40	C	T	C	A	G	C	C	C	A	C	C
## 41	C	T	C	A	G	C	C	C	A	C	C
##	loc-46	loc-47	loc-48	loc-49	loc-50	loc-51	loc-52	loc-53	loc-54	loc-55	loc-56
## 1	A	C	C	A	G	T	G	T	T	T	G
## 2	A	C	C	A	G	T	G	T	T	T	G
## 3	A	C	C	A	G	T	G	T	T	T	G
## 4	A	C	C	A	G	T	G	T	T	T	G
## 5	A	C	C	A	G	T	G	T	T	T	G
## 6	A	C	C	A	G	T	G	T	T	T	G
## 7	A	T	C	G	A	T	G	G	T	T	G
## 8	A	T	C	A	A	T	G	G	T	T	G
## 9	A	T	C	A	A	T	G	G	T	T	G
## 10	A	T	C	A	A	T	G	G	T	T	G
## 11	C	T	C	A	A	T	G	G	T	T	G
## 12	C	T	C	A	A	T	G	G	T	T	G
## 13	C	T	C	A	A	T	G	G	T	T	G
## 14	C	T	C	A	A	T	G	G	T	T	G
## 15	A	T	C	A	A	T	G	G	T	T	G
## 16	A	T	C	G	A	T	G	G	T	T	G
## 17	A	T	C	G	A	T	G	G	T	T	G
## 18	A	C	C	A	G	T	G	T	T	T	G
## 19	A	C	C	A	G	T	G	T	T	T	G
## 20	A	C	C	A	G	T	G	T	T	T	G
## 21	A	C	C	A	G	T	G	T	T	T	G
## 22	A	C	C	A	G	T	G	T	T	T	G
## 23	A	C	C	A	G	T	G	T	T	T	G
## 24	A	C	C	A	G	T	G	T	T	T	G
## 25	A	C	C	A	G	T	G	T	T	T	G
## 26	A	C	C	A	G	T	G	T	T	T	G
## 27	A	C	C	A	G	T	G	T	T	T	G

## 28	A	C	C	A	G	T	G	T	T	T	G
## 29	A	C	C	A	G	T	G	T	T	T	G
## 30	A	C	C	A	G	T	G	T	T	T	G
## 31	A	C	C	A	G	T	G	T	T	T	G
## 32	A	C	C	A	G	T	G	T	T	T	G
## 33	A	C	C	A	G	T	G	T	T	T	G
## 34	A	C	C	A	G	T	G	T	T	T	G
## 35	A	C	C	A	G	T	G	T	T	T	G
## 36	A	C	C	A	G	T	G	T	T	T	G
## 37	A	T	C	A	A	T	G	G	T	T	G
## 38	A	C	C	A	G	T	G	T	T	T	G
## 39	A	C	C	A	G	T	G	T	T	T	G
## 40	A	C	C	A	G	T	G	T	T	T	G
## 41	A	C	C	A	G	T	G	T	T	T	G
##	loc-57	loc-58	loc-59	loc-60	loc-61	loc-62	loc-63	loc-64	loc-65	loc-66	loc-67
## 1	C	C	C	C	T	G	C	A	C	A	G
## 2	C	C	C	C	T	G	C	A	C	A	G
## 3	C	C	C	C	T	G	C	A	C	A	G
## 4	C	C	C	C	T	G	C	A	C	A	G
## 5	C	C	C	C	T	G	C	A	C	A	G
## 6	C	C	C	C	T	G	C	A	C	A	G
## 7	C	C	C	C	T	G	C	A	C	A	G
## 8	C	C	C	C	T	G	C	A	C	A	G
## 9	C	C	C	C	T	G	C	A	C	A	G
## 10	C	C	C	C	T	G	C	A	C	A	G
## 11	C	C	C	C	T	G	C	A	C	A	G
## 12	C	C	C	C	T	G	C	A	C	A	G
## 13	C	C	T	C	T	G	C	A	C	A	G
## 14	C	C	T	C	T	G	C	A	C	A	G
## 15	C	C	C	C	T	G	C	A	C	A	G
## 16	C	C	C	C	T	G	C	A	C	A	G
## 17	C	C	C	C	T	G	C	A	C	A	G
## 18	C	C	C	C	T	G	C	A	C	A	G
## 19	C	C	C	C	T	G	C	A	C	A	G
## 20	C	C	C	C	T	G	C	A	C	A	G
## 21	C	C	C	C	T	G	C	A	C	A	G
## 22	C	C	C	C	T	G	C	A	C	A	G
## 23	C	C	C	C	T	G	C	A	C	A	G
## 24	C	C	C	C	T	G	C	A	C	A	G
## 25	C	C	C	C	T	G	C	A	C	A	G
## 26	C	C	C	C	T	G	C	A	C	A	G
## 27	C	C	C	C	T	G	C	A	C	A	G
## 28	C	C	C	C	T	G	C	A	C	A	G
## 29	C	C	C	C	T	G	C	A	C	A	G
## 30	C	C	C	C	T	G	C	A	C	A	G
## 31	C	C	C	C	T	G	C	A	C	A	G
## 32	C	C	C	C	T	G	C	A	C	A	G
## 33	C	C	C	C	T	G	C	A	C	A	G
## 34	C	C	C	C	T	G	C	A	C	A	G
## 35	C	C	C	C	T	G	C	A	C	A	G
## 36	C	C	C	C	T	G	C	A	C	A	G
## 37	C	C	C	C	T	G	C	A	C	A	G
## 38	C	C	C	C	T	G	C	A	C	A	G
## 39	C	C	C	C	T	G	C	A	C	A	G

## 40	C	C	C	C	T	G	C	A	C	A	G
## 41	C	C	C	C	T	G	C	A	C	A	G
##	loc-68	loc-69	loc-70	loc-71	loc-72	loc-73	loc-74	loc-75	loc-76	loc-77	loc-78
## 1	G	G	G	C	A	G	C	G	C	C	T
## 2	G	G	G	C	A	G	C	G	C	A	T
## 3	G	G	G	C	A	G	C	G	C	C	T
## 4	G	G	G	C	A	G	C	G	C	C	T
## 5	G	G	G	C	A	G	C	G	C	C	T
## 6	G	G	G	C	A	G	C	G	C	C	T
## 7	G	G	G	C	A	G	C	G	C	C	C
## 8	G	G	G	C	A	G	C	G	C	C	C
## 9	G	G	G	C	A	G	C	G	C	C	C
## 10	G	G	G	C	A	G	C	G	C	C	C
## 11	G	G	G	C	A	G	C	G	C	C	T
## 12	G	G	G	C	A	G	C	G	C	C	T
## 13	G	G	G	C	A	G	C	G	C	C	T
## 14	G	G	G	C	A	G	C	G	C	C	T
## 15	G	G	G	C	A	G	C	G	C	C	T
## 16	G	G	G	C	A	G	C	G	C	C	C
## 17	G	G	G	C	A	G	C	G	C	C	C
## 18	G	G	G	C	A	G	C	G	C	C	C
## 19	G	G	G	C	A	G	C	G	C	C	C
## 20	G	G	G	C	A	G	C	G	C	C	C
## 21	G	G	G	C	A	G	C	G	C	C	C
## 22	G	G	G	C	A	G	C	G	C	C	C
## 23	G	G	G	C	A	G	C	G	C	C	C
## 24	G	G	G	C	A	G	C	G	C	C	C
## 25	G	G	G	C	A	G	C	G	C	C	C
## 26	G	G	G	C	A	G	C	G	C	C	T
## 27	G	G	G	C	A	G	C	G	C	C	T
## 28	G	G	G	C	A	G	C	G	C	C	T
## 29	G	G	G	C	A	G	C	G	C	C	T
## 30	G	G	G	C	A	G	C	G	C	C	T
## 31	G	G	G	C	A	G	C	G	C	C	T
## 32	G	G	G	C	A	G	C	G	C	C	T
## 33	G	G	G	C	A	G	C	G	C	C	T
## 34	G	G	G	C	A	G	C	G	C	C	T
## 35	G	G	G	C	A	G	C	G	C	C	C
## 36	G	G	G	C	A	G	C	G	C	C	T
## 37	G	G	G	C	A	G	C	G	C	C	C
## 38	G	G	G	C	A	G	C	G	C	C	C
## 39	G	G	G	C	A	G	C	G	C	C	T
## 40	G	G	G	C	A	G	C	G	C	C	T
## 41	G	G	G	C	A	G	C	G	C	C	T
##	loc-79	loc-80	loc-81	loc-82	loc-83	loc-84	loc-85	loc-86	loc-87	loc-88	loc-89
## 1	A	T	C	G	G	T	C	A	T	G	A
## 2	A	T	C	G	G	T	C	A	T	G	A
## 3	A	T	C	G	G	T	C	A	T	G	A
## 4	A	T	C	G	G	T	C	A	T	G	A
## 5	A	T	C	G	G	T	C	A	T	G	A
## 6	A	T	C	G	G	T	C	A	T	G	A
## 7	A	T	T	G	G	A	C	A	T	G	A
## 8	A	T	T	G	G	T	C	A	T	G	A
## 9	A	T	T	G	G	T	C	A	T	G	A

## 10	A	T	T	G	G	T	C	A	T	G	A
## 11	T	T	T	G	G	T	C	A	T	G	A
## 12	T	T	T	G	G	T	C	A	T	G	A
## 13	T	T	T	G	G	T	C	A	T	G	A
## 14	T	T	T	G	G	T	C	A	T	G	A
## 15	A	T	T	G	G	T	C	A	T	G	A
## 16	A	T	T	G	G	A	C	A	T	G	A
## 17	A	T	T	G	G	A	C	A	T	G	A
## 18	A	T	C	G	G	T	C	A	T	G	A
## 19	A	T	C	G	G	T	C	A	T	G	A
## 20	A	T	C	G	G	T	C	A	T	G	A
## 21	A	T	C	G	G	T	C	A	T	G	A
## 22	A	T	C	G	G	T	C	A	T	G	A
## 23	A	T	C	G	G	T	C	A	T	G	A
## 24	A	T	C	G	G	T	C	A	T	G	A
## 25	T	T	C	G	G	T	C	A	T	G	A
## 26	A	T	C	G	G	T	C	A	T	G	A
## 27	A	T	C	G	G	T	C	A	T	G	A
## 28	A	T	C	G	G	T	C	A	T	G	A
## 29	A	T	C	G	G	T	C	A	T	G	A
## 30	A	T	C	G	G	T	C	A	T	G	A
## 31	A	T	C	G	G	T	C	A	T	G	A
## 32	A	T	C	G	G	T	C	A	T	G	A
## 33	A	T	C	G	G	T	C	A	T	G	A
## 34	A	T	C	G	G	T	C	A	T	G	A
## 35	A	T	C	G	G	T	C	A	T	G	A
## 36	A	T	C	G	G	T	C	A	T	G	A
## 37	A	T	T	G	G	T	C	A	T	G	A
## 38	A	T	C	G	G	T	C	A	T	G	A
## 39	A	T	C	G	G	T	C	A	T	G	A
## 40	A	T	C	G	G	T	C	A	T	G	A
## 41	A	T	C	G	G	T	C	A	T	G	A
##	loc-90	loc-91	loc-92	loc-93	loc-94	loc-95	loc-96	loc-97	loc-98	loc-99	
## 1	C	C	C	G	G	T	G	G	T	A	
## 2	C	C	C	G	G	T	G	G	T	A	
## 3	C	C	C	G	G	T	A	G	T	A	
## 4	C	C	C	G	G	T	G	G	A	A	
## 5	C	C	C	G	G	T	G	G	T	A	
## 6	C	C	C	G	G	T	G	G	T	A	
## 7	C	C	C	G	G	T	G	G	T	A	
## 8	C	C	C	G	G	T	G	G	T	A	
## 9	C	C	C	G	G	T	G	G	T	A	
## 10	C	C	C	G	G	T	G	G	T	A	
## 11	C	C	C	G	G	T	G	G	T	A	
## 12	C	C	C	G	G	T	G	G	T	A	
## 13	C	C	C	G	G	T	G	G	T	A	
## 14	C	C	C	G	G	T	G	G	T	A	
## 15	C	C	C	G	G	T	G	G	T	A	
## 16	C	C	C	G	G	T	G	G	T	A	
## 17	C	C	C	G	G	T	G	G	T	A	
## 18	C	C	C	G	G	T	A	G	T	A	
## 19	C	C	C	G	G	T	G	A	T	A	
## 20	C	C	C	G	G	T	G	G	A	A	
## 21	C	C	C	G	G	T	G	G	A	A	

## 22	C	C	C	G	G	T	G	G	T	A
## 23	C	C	C	G	G	T	G	G	T	A
## 24	C	C	C	G	G	T	G	G	T	A
## 25	C	C	C	G	G	T	G	G	T	A
## 26	C	C	C	G	G	T	A	G	T	A
## 27	C	C	C	G	G	T	G	G	A	A
## 28	C	C	C	G	G	T	G	G	A	A
## 29	C	C	C	G	G	T	G	G	T	A
## 30	C	C	C	G	G	T	G	G	T	A
## 31	C	C	C	G	G	T	G	G	T	A
## 32	C	C	C	G	G	T	A	G	T	A
## 33	C	C	C	G	G	T	G	G	T	A
## 34	C	C	C	G	G	T	G	G	T	A
## 35	C	C	C	G	G	T	A	G	A	A
## 36	C	C	C	G	G	T	G	G	T	A
## 37	C	C	C	G	G	T	G	G	T	A
## 38	C	C	C	G	G	T	G	G	T	A
## 39	C	C	C	G	G	T	G	G	A	A
## 40	C	C	C	G	G	T	G	G	T	A
## 41	C	C	C	G	G	T	G	G	T	A
##	loc-100	loc-101	loc-102	loc-103	loc-104	loc-105	loc-106	loc-107	loc-108	
## 1	T	T	G	A	A	A	A	T	T	
## 2	T	T	G	A	A	A	A	T	T	
## 3	T	T	G	A	A	A	A	T	T	
## 4	T	T	G	A	A	A	A	T	T	
## 5	T	T	G	A	A	A	A	T	T	
## 6	T	T	G	G	A	A	A	T	T	
## 7	T	T	G	A	A	A	A	T	G	
## 8	T	T	G	A	A	A	A	T	G	
## 9	T	T	G	A	A	A	A	T	G	
## 10	T	T	G	G	A	A	A	T	G	
## 11	T	T	G	A	A	A	A	T	G	
## 12	T	T	G	G	A	A	A	T	G	
## 13	T	T	G	A	A	A	A	T	G	
## 14	T	T	G	A	G	A	A	T	G	
## 15	T	T	G	A	A	A	A	T	G	
## 16	T	T	G	A	A	A	A	T	G	
## 17	T	T	G	A	A	A	A	T	G	
## 18	T	T	G	A	A	A	A	T	T	
## 19	T	T	G	A	A	A	A	T	T	
## 20	T	T	G	A	A	A	A	T	T	
## 21	T	T	G	G	A	A	A	T	T	
## 22	T	T	G	A	A	A	A	T	T	
## 23	T	T	G	A	A	A	A	T	T	
## 24	T	T	G	G	A	A	A	T	T	
## 25	T	T	G	A	A	A	A	T	T	
## 26	T	T	G	A	A	A	A	T	T	
## 27	T	T	G	A	A	A	A	T	T	
## 28	T	T	G	G	A	A	A	T	T	
## 29	T	T	G	A	A	A	A	T	G	
## 30	T	T	G	A	A	A	A	T	T	
## 31	T	T	G	G	A	A	A	T	T	
## 32	T	T	G	A	A	A	A	T	T	
## 33	T	T	G	A	A	A	A	T	T	

## 34	T	T	G	G	A	A	A	T	T
## 35	T	T	G	A	A	A	A	T	T
## 36	T	T	G	A	A	A	A	T	T
## 37	T	T	G	A	A	A	A	T	G
## 38	T	T	G	A	A	A	A	T	T
## 39	T	T	G	A	A	A	A	T	T
## 40	T	T	G	A	A	A	A	T	G
## 41	T	T	G	A	A	A	A	T	T
##	loc-109	loc-110	loc-111	loc-112	loc-113	loc-114	loc-115	loc-116	loc-117
## 1	A	G	T	T	A	G	C	G	A
## 2	A	G	T	T	A	G	C	G	A
## 3	A	G	T	T	A	G	C	G	A
## 4	A	G	T	T	A	G	C	G	A
## 5	A	G	T	T	A	G	C	G	A
## 6	A	G	T	T	A	G	C	G	A
## 7	A	G	T	T	G	T	A	G	A
## 8	A	G	T	T	A	T	A	G	A
## 9	A	G	A	T	A	T	A	G	A
## 10	A	G	A	T	A	T	A	G	A
## 11	A	G	T	T	A	T	C	G	A
## 12	A	G	T	T	A	T	C	G	A
## 13	A	G	T	T	A	T	C	G	A
## 14	A	G	T	T	A	T	C	G	A
## 15	A	G	T	T	A	T	A	G	A
## 16	A	G	T	T	G	T	A	G	A
## 17	A	G	T	T	G	T	A	G	A
## 18	A	G	T	T	A	G	C	G	A
## 19	A	G	T	T	A	G	C	G	A
## 20	A	G	T	T	A	G	C	G	A
## 21	A	G	T	T	A	G	C	G	A
## 22	A	G	T	T	A	G	C	G	A
## 23	A	G	T	T	A	G	C	G	A
## 24	A	G	T	T	A	G	C	G	A
## 25	A	G	T	T	A	G	C	G	A
## 26	A	G	T	T	A	G	C	G	A
## 27	A	G	T	T	A	G	C	G	A
## 28	A	G	T	T	A	G	C	G	A
## 29	A	G	T	T	G	T	A	G	A
## 30	A	G	T	T	A	G	C	G	A
## 31	A	G	T	T	A	G	C	G	A
## 32	A	G	T	T	A	G	C	G	A
## 33	A	G	T	T	A	G	C	G	A
## 34	A	G	T	T	A	G	C	G	A
## 35	A	G	T	T	A	G	C	G	A
## 36	A	G	T	T	A	G	C	G	A
## 37	A	G	A	T	A	T	A	G	A
## 38	A	G	T	T	A	G	C	G	A
## 39	A	G	T	T	A	G	C	G	A
## 40	A	G	T	T	G	T	A	G	A
## 41	A	G	T	T	A	G	C	G	A
##	loc-118	loc-119	loc-120	loc-121	loc-122	loc-123	loc-124	loc-125	loc-126
## 1	C	G	G	G	G	G	T	T	C
## 2	C	G	G	G	G	G	T	T	C
## 3	C	G	G	G	G	G	T	T	C

## 4	C	G	G	G	G	G	T	T	C
## 5	C	G	G	G	G	G	T	T	C
## 6	C	G	G	G	G	G	T	T	C
## 7	C	G	G	G	G	G	C	T	C
## 8	C	G	G	G	G	G	C	T	C
## 9	C	G	G	G	G	G	C	T	C
## 10	C	G	G	G	G	G	C	T	C
## 11	C	G	G	G	G	G	C	T	C
## 12	C	G	G	G	G	G	C	T	C
## 13	C	G	G	G	G	G	C	T	C
## 14	C	G	G	G	G	G	C	T	C
## 15	C	G	G	G	G	G	C	T	C
## 16	C	G	G	G	G	G	C	T	C
## 17	C	G	G	G	G	G	C	T	C
## 18	C	G	G	G	G	G	T	T	C
## 19	C	G	G	G	G	G	T	T	C
## 20	C	G	G	G	G	G	T	T	C
## 21	C	G	G	G	G	G	T	T	C
## 22	C	G	G	G	G	G	T	T	C
## 23	C	G	G	G	G	G	T	T	C
## 24	C	G	G	G	G	G	T	T	C
## 25	C	G	G	G	G	G	T	T	C
## 26	C	G	G	G	G	G	T	T	C
## 27	C	G	G	G	G	G	T	T	C
## 28	C	G	G	G	G	G	T	T	C
## 29	C	G	G	G	G	G	C	T	C
## 30	C	G	G	G	G	G	T	T	C
## 31	C	G	G	G	G	G	T	T	C
## 32	C	G	G	G	G	G	T	T	C
## 33	C	G	G	G	G	G	T	T	C
## 34	C	G	G	G	G	G	T	T	C
## 35	C	G	G	G	G	G	T	T	C
## 36	C	G	G	G	G	G	T	T	C
## 37	C	G	G	G	G	G	C	T	C
## 38	C	G	G	G	G	G	T	T	C
## 39	C	G	G	G	G	G	T	T	C
## 40	C	G	G	G	G	G	C	T	C
## 41	C	G	G	G	G	G	T	T	C
##	loc-127	loc-128	loc-129	loc-130	loc-131	loc-132	loc-133	loc-134	loc-135
## 1	T	T	C	C	G	A	A	C	A
## 2	T	T	C	C	G	A	A	C	A
## 3	T	T	C	C	G	A	A	C	A
## 4	T	T	C	C	G	A	A	C	A
## 5	T	T	C	C	G	A	A	C	A
## 6	T	T	C	C	G	A	A	C	A
## 7	T	T	C	C	G	A	A	A	A
## 8	T	T	C	C	G	A	A	A	A
## 9	T	T	C	C	G	A	A	A	A
## 10	T	T	C	C	G	A	A	A	A
## 11	T	T	C	C	G	A	A	A	A
## 12	T	T	C	C	G	A	A	A	A
## 13	T	T	C	C	G	A	A	A	A
## 14	T	T	C	C	G	A	A	A	A
## 15	T	T	C	C	G	A	A	A	A

## 16	T	T	C	C	G	A	A	A	A
## 17	T	T	C	C	G	A	A	A	A
## 18	T	T	C	C	G	A	A	C	A
## 19	T	T	C	C	G	A	A	C	A
## 20	T	T	C	C	G	A	A	C	A
## 21	T	T	C	C	G	A	A	C	A
## 22	T	T	C	C	G	A	A	A	A
## 23	T	T	C	C	G	A	A	C	A
## 24	T	T	C	C	G	A	A	C	A
## 25	T	T	C	C	G	A	A	C	A
## 26	T	T	C	C	G	A	A	C	A
## 27	T	T	C	C	G	A	A	C	A
## 28	T	T	C	C	G	A	A	C	A
## 29	T	T	C	C	G	A	A	A	A
## 30	T	T	C	C	G	A	A	C	A
## 31	T	T	C	C	G	A	A	C	A
## 32	T	T	C	C	G	A	A	C	A
## 33	T	T	C	C	G	A	A	C	A
## 34	T	T	C	C	G	A	A	C	A
## 35	T	T	C	C	G	A	A	C	A
## 36	T	T	C	C	G	A	A	C	A
## 37	T	T	C	C	G	A	A	A	A
## 38	T	T	C	C	G	A	A	C	A
## 39	T	T	C	C	G	A	A	C	A
## 40	T	T	C	C	G	A	A	A	A
## 41	T	T	C	C	G	A	A	C	A
##	loc-136	loc-137	loc-138	loc-139	loc-140	loc-141	loc-142	loc-143	loc-144
## 1	C	T	T	C	C	C	C	G	T
## 2	C	T	T	C	C	C	C	G	T
## 3	C	T	T	C	C	C	C	G	T
## 4	C	T	T	C	C	C	C	G	T
## 5	C	T	T	C	C	C	C	G	T
## 6	C	T	T	C	C	C	C	G	T
## 7	C	T	T	C	C	C	C	G	T
## 8	C	T	T	C	C	C	C	G	T
## 9	C	T	T	C	C	C	C	G	T
## 10	C	T	T	C	C	C	C	G	T
## 11	G	T	T	C	C	C	C	G	T
## 12	G	T	T	C	C	C	C	G	T
## 13	G	T	T	C	C	C	C	G	T
## 14	G	T	T	C	C	C	C	G	T
## 15	C	T	T	C	C	C	C	G	T
## 16	C	T	T	C	C	C	C	G	T
## 17	C	T	T	C	C	C	C	G	T
## 18	C	T	T	C	C	C	C	G	T
## 19	C	T	T	C	C	C	C	G	T
## 20	C	T	T	C	C	C	C	G	T
## 21	C	T	T	C	C	C	C	G	T
## 22	C	T	T	C	C	C	C	G	T
## 23	C	T	T	C	C	C	C	G	T
## 24	C	T	T	C	C	C	C	G	T
## 25	C	T	T	C	C	C	C	G	T
## 26	C	T	T	C	C	C	C	G	T
## 27	C	T	T	C	C	C	C	G	T

## 28	C	T	T	C	C	C	C	G	T
## 29	C	T	T	C	C	C	C	G	T
## 30	C	T	T	C	C	C	C	G	T
## 31	C	T	T	C	C	C	C	G	T
## 32	C	T	T	C	C	C	C	G	T
## 33	C	T	T	C	C	C	C	G	T
## 34	C	T	T	C	C	C	C	G	T
## 35	C	T	T	C	C	C	C	G	T
## 36	C	T	T	C	C	C	C	G	T
## 37	C	T	T	C	C	C	C	G	T
## 38	C	T	T	C	C	C	C	G	T
## 39	C	T	T	C	C	C	C	G	T
## 40	C	T	T	C	C	C	C	G	T
## 41	C	T	T	C	C	C	C	G	T
##	loc-145	loc-146	loc-147	loc-148	loc-149	loc-150	loc-151	loc-152	loc-153
## 1	C	A	A	C	C	C	C	A	C
## 2	C	A	A	C	C	C	C	A	C
## 3	C	A	A	C	C	C	C	A	C
## 4	C	A	A	C	C	C	C	A	C
## 5	C	A	A	C	C	C	C	A	C
## 6	C	A	A	C	C	C	C	A	C
## 7	C	A	A	C	C	C	C	A	C
## 8	C	A	A	C	C	C	C	A	C
## 9	C	A	A	C	C	C	C	A	C
## 10	C	A	A	C	C	C	C	A	C
## 11	C	A	A	C	C	C	C	A	C
## 12	C	A	A	C	C	C	C	A	C
## 13	C	A	A	C	C	C	C	A	C
## 14	C	A	A	C	C	C	C	A	C
## 15	C	A	A	C	C	C	C	A	C
## 16	C	A	A	C	C	C	C	A	C
## 17	C	A	A	C	C	C	C	A	C
## 18	C	A	A	C	C	C	C	A	C
## 19	C	A	A	C	C	C	C	A	C
## 20	C	A	A	C	C	C	C	A	C
## 21	C	A	A	C	C	C	C	A	C
## 22	C	A	A	C	C	C	C	A	C
## 23	C	A	A	C	C	C	C	A	C
## 24	C	A	A	C	C	C	C	A	C
## 25	C	A	A	C	C	C	C	A	C
## 26	C	A	A	C	C	C	C	A	C
## 27	C	A	A	C	C	C	C	A	C
## 28	C	A	A	C	C	C	C	A	C
## 29	C	A	A	C	C	C	C	A	C
## 30	C	A	A	C	C	C	C	A	C
## 31	C	A	A	C	C	C	C	A	C
## 32	C	A	A	C	C	C	C	A	C
## 33	C	A	A	C	C	C	C	A	C
## 34	C	A	A	C	C	C	C	A	C
## 35	C	A	A	C	C	C	C	A	C
## 36	C	A	A	C	C	C	C	A	C
## 37	C	A	A	C	C	C	C	A	C
## 38	C	A	A	C	C	C	C	A	C
## 39	C	A	A	C	C	C	C	A	C

	loc-154	loc-155	loc-156	loc-157	loc-158	loc-159	loc-160	loc-161	loc-162
## 40	C	A	A	C	C	C	C	A	C
## 41	C	A	A	C	C	C	C	A	C
## 1	A	G	G	C	G	C	G	T	A
## 2	A	G	G	C	G	C	G	T	A
## 3	A	G	G	C	G	C	G	T	A
## 4	A	G	G	C	G	C	G	T	A
## 5	A	G	G	C	G	C	G	T	A
## 6	A	G	G	C	G	C	G	T	A
## 7	A	G	G	C	G	T	G	C	T
## 8	A	G	G	C	G	C	G	C	T
## 9	A	G	G	C	G	C	G	C	T
## 10	A	G	G	C	G	C	G	C	T
## 11	A	G	G	C	G	C	G	C	T
## 12	A	G	G	C	G	C	G	C	T
## 13	A	G	G	C	G	C	G	C	T
## 14	A	G	G	C	G	C	G	C	T
## 15	A	G	G	C	G	C	G	C	T
## 16	A	G	G	C	G	T	G	C	T
## 17	A	G	G	C	G	T	G	C	T
## 18	A	G	G	C	G	C	G	T	A
## 19	A	G	G	C	G	C	G	T	A
## 20	A	G	G	C	G	C	G	T	A
## 21	A	G	G	C	G	C	G	T	A
## 22	A	G	G	C	G	C	G	T	A
## 23	A	G	G	C	G	C	G	T	A
## 24	A	G	G	C	G	C	G	T	A
## 25	A	G	G	C	G	C	G	T	A
## 26	A	G	G	C	G	C	G	T	A
## 27	A	G	G	C	G	C	G	T	A
## 28	A	G	G	C	G	C	G	T	A
## 29	A	G	G	C	G	T	G	C	T
## 30	A	G	G	C	G	C	G	T	A
## 31	A	G	G	C	G	C	G	T	A
## 32	A	G	G	C	G	C	G	T	A
## 33	A	G	G	C	G	C	G	T	A
## 34	A	G	G	C	G	C	G	T	A
## 35	A	G	G	C	G	C	G	T	A
## 36	A	G	G	C	G	C	G	T	A
## 37	A	G	G	C	G	C	G	C	T
## 38	A	G	G	C	G	C	G	T	A
## 39	A	G	G	C	G	C	G	T	A
## 40	A	G	G	C	G	T	G	C	T
## 41	A	G	G	C	G	C	G	T	A
##	loc-163	hap.freq							
## 1	G	0.00467							
## 2	G	0.00467							
## 3	G	0.00422							
## 4	G	0.00344							
## 5	G	0.04012							
## 6	G	0.00050							
## 7	A	0.00467							
## 8	A	0.00467							
## 9	A	0.06325							

```

## 10      A  0.00041
## 11      A  0.01869
## 12      A  0.00467
## 13      A  0.04673
## 14      A  0.00467
## 15      A  0.00467
## 16      A  0.01869
## 17      A  0.00467
## 18      G  0.00913
## 19      G  0.00467
## 20      G  0.04632
## 21      G  0.00435
## 22      G  0.00467
## 23      G  0.01588
## 24      G  0.00758
## 25      G  0.00467
## 26      G  0.00534
## 27      G  0.00400
## 28      G  0.00378
## 29      G  0.00760
## 30      G  0.05770
## 31      G  0.04879
## 32      G  0.00467
## 33      G  0.02804
## 34      G  0.00467
## 35      G  0.00467
## 36      G  0.03116
## 37      A  0.06718
## 38      G  0.01487
## 39      G  0.00820
## 40      G  0.01109
## 41      G  0.36750
## =====
##                                     Details
## =====
## lnlike =  -472.8509
## lr stat for no LD =  2838.267 , df =  -3 , p-val =  NA

```

There are 163 haplotypes possibles. The most frequent is number 6.

```

bim_file <- "APOE/APOE.bim"
bim_data <- read.table(bim_file, header = FALSE, stringsAsFactors = FALSE)

genotype_matrix <- data[,-1]
for (col in colnames(genotype_matrix)) {
  genotype_matrix[[col]] <- genotype(genotype_matrix[[col]])
}
calculate_MAF <- function(genotype_counts) {
  allele1 <- 2 * genotype_counts[[1]] + genotype_counts[[2]]
  allele2 <- 2 * genotype_counts[[3]] + genotype_counts[[2]]
  return (min(allele1, allele2) / (allele2 + allele1))
}

```

```

filter_index_MAF <- c()
res <- 0
for (i in 1:ncol(genotype_matrix)) {
  alleles <- c(paste(bim_data$V5[[i]], "/", bim_data$V6[[i]], sep = ""))
  count_gentoype <- MakeCounts(genotype_matrix[[i]], alleles = alleles, sep="/")
  MAF <- calculate_MAF(count_gentoype)
  if (MAF < 0.1) {
    filter_index_MAF <- c(filter_index_MAF, i)
  }
}
filter_index_MAF <- filter_index_MAF + 1
tmp1 <- 2 * filter_index_MAF
tmp2 <- 2 * filter_index_MAF + 1
filter_index_MAF <- cbind(tmp1, tmp2)
Geno_filtered <- Geno[, -filter_index_MAF]
print(haplo.em(Geno_filtered))

```

Question 4

```

## =====
##                                     Haplotypes
## =====
##   loc-1 loc-2 loc-3 loc-4 loc-5 loc-6 loc-7 loc-8 loc-9 loc-10 loc-11 loc-12
## 1      2      C      A      C      C      G      T      C      T      A      G      C
## 2      2      C      G      C      C      G      T      C      T      A      G      T
## 3      2      G      G      T      T      A      C      A      C      G      T      C
## 4      2      G      G      T      T      A      C      A      C      G      T      T
## 5      2      G      G      T      T      A      C      A      C      G      T      T
## 6      N      C      C      G      A      C      G      C      C      C      T      T
## 7      N      C      C      G      C      C      G      C      C      C      T      C
## 8      N      C      C      G      C      C      G      C      C      C      T      C
## 9      N      C      C      G      C      C      G      C      C      C      T      T
## 10     N      C      C      G      C      C      G      C      C      C      T      T
## 11     N      C      C      G      C      C      G      C      C      C      T      T
## 12     N      C      C      G      C      C      G      C      C      C      T      T
## 13     N      G      C      G      C      C      G      C      C      C      T      T
##   loc-13 loc-14 loc-15 loc-16 loc-17 loc-18 loc-19 loc-20 loc-21 loc-22
## 1      T      G      A      T      A      C      A      C      T      A
## 2      T      G      T      T      A      C      A      C      T      A
## 3      C      T      T      G      A      T      C      T      T      A
## 4      C      G      T      T      A      C      A      C      T      G
## 5      C      T      T      G      C      T      C      T      A      G
## 6      A      G      A      T      C      G      T      A      A      G
## 7      A      G      A      T      A      G      T      A      T      A
## 8      T      G      A      T      A      G      T      A      T      A
## 9      A      G      A      T      A      G      T      A      T      A
## 10     A      G      A      T      A      G      T      A      T      G
## 11     A      G      A      T      C      G      T      A      A      G
## 12     T      G      A      T      C      G      T      A      T      A
## 13     A      G      A      T      C      G      T      A      A      G
##   hap.freq
## 1      0.00467
## 2      0.02336
## 3      0.00000

```

```

## 4    0.00935
## 5    0.46262
## 6    0.00467
## 7    0.02336
## 8    0.00467
## 9    0.11215
## 10   0.00935
## 11   0.26636
## 12   0.07477
## 13   0.00467
## =====
##                                     Details
## =====
## lnlike =  -249.2797
## lr stat for no LD =  752.5871 , df =  -21 , p-val =  NA

```

There are only 22 haplotypes and the most frequent is number 5.

Excluding genetic variants with a minor allele frequency (MAF) below 0.10 led to a significant reduction in the number of identified haplotypes, from 163 to 22. This reduction reflects a focus on more common variants in the population, excluding those with lower frequencies. The decrease suggests a loss of genetic diversity, as the excluded haplotypes were likely associated with less common variants in the studied population. The choice of the MAF threshold is crucial and may impact haplotype analysis outcomes, highlighting the need for careful consideration based on the study's objectives.