Q1)

a)

Q

results as expected?

State assumptions

METHOD

* calculate the heterozygosity for each of the columns (SNPs) for one of the population
* Then find the median using all SNPs
* repeat the above stages to calculate for each of the populations
* Calculate as a ratio
* Calculation:

ANS =

* Median heterzygosity
* Ceu – 0.299
* Jpt – 0.279
* Tsi – 0.307
* Yri – 0.304
* PopX – 0.311

Resources – (to get the sum of every two rows in R; stack overflow)

ANS=

b)

METHOD –

Calculate the Fst, between every pairing, for each column (SNP)

Then find the median using all SNPs of that population

Repeat for each of the populations

Should output 5 medians

Use of a loop for each of the colums?

The median rm code should remove the NA for each of the SNP colums

does the phrase between “every pairing” make a difference?

Check fst info from previous exercises

Can check the previous assignment for similarities of differences

expected, why, why not?

c)

where do you think PopX comes from?

ANS =

could Popx be a combination of the other populations?

d)

first five columns (SNPs); first 4 rows (first two individuals)

Buld an Ancestral Recombination Graph (ARG) 5 SNP region

METHOD - research Ancestral Recombination Graph (ARG)

A close up of text on a white background

Description automatically generated

A close up of text on a white background

Description automatically generated

e)

Arg as expected, why or why not?

Ten sequences, 5 diploid, time until each of the coalescent events?

All individuals coalesce?

METHOD - research time to coalesce (each / all) ?

**Q2)**

**EXPLANATION**

Two biallelic loci

Init (vector) – relative frequency of the four haplotypes {AB, Ab, aB, ab} (gen 1)

Population size is diploid (x 2 for haploid)

Nall – the total number of diploid individuals

Nall default = 2000

Haplotypes = 4000

The A allelic type at first locus is undergoing selection

Parameter S controlling the increase in fitness

S – selection coefficient

RHO – rate of recombination between the two loci

to study recombination we need to consider diploid

individuals and so nall here refers to the total number of diploid individuals. The

default is nall = 2000 individuals, which corresponds to 4000 haplotypes.

**OUTPUT** – vector 6 elements

Ngen default 500 generations

First 4 elements haplotype proportions {AB, Ab, aB, ab}

Last 2 elements |D’| and r2 (between the two loci)

DAB = 0; otherwise A;B are in Linkage

Disequilibrium (LD).

YES – two plots

1. One gives the **haplotype frequency** trajectories over time

And the frequency trajectories of the **allele frequencies** of the allele frequencies A and B

1. Linkage disequilibrium measures **r2**and **|D’|** between the two loci over this same time frame

Only simulate one population at a time so can illustrate the frequencies of different haplotypes (many lines for one population instead of a single line for each population

NO – to.plot = no

Use default values nall = 2000, mu=0

Note that unlike with wf.R, you can only simulate one population at a

time with ldsel, so that the plot can illustrate the frequencies of the di  
erent hapdlotypes.

Advantage: jD0j = 1 means no evidence of recombination between the markers under infinite-sites (only three of four possible haplotypes are present)

D0 and r2 low

**QUESTIONS**

a)

METHOD –

* (i.e. npop x init / nall).

ANS=

- A allele equal to AB and Ab?

- or A on the allele frequencies graph?

b)

add recombination (rho)

how change?

Figure labelling e.g. Measures of LD in a simulation of the 2-locus WF model, with u = 0

and p = 0.001.

ANS =

For recombination the corresponding measures of I D’ I and r2 should have low values indicating that recombination is present.

In general: I D’ I and r 2 usually decrease with increasing recombination, but are affected by loci's allele frequencies

c)

new mutation A allele

enters the population and undergoes selection (advantageous for survival)

**rho = 0 (rate of recombination)**

METHOD:

* Change the initial values so that the selected allele A has an initial frequency 0.01 ??
* Varying s (selection) what is the pattern over time
* For one value of S find the median time to fixation of the A allele

ANS =

New mutants move I D’ I towards 1, then decline towards 0

r2 also **increases** but less so, depending on frequency of haplotype on

which new mutant arises

If the novel mutant **increases in frequency**, I D’ I will quickly **decline** to 0 unless p (rho) small

If p (rho) small, I D’ I can remain high for many generations, and meanwhile r2 can also become large

If the founding **mutation** events occurred on the **same branch** of the coalescent tree (occurred at a similar time).

**r 2 = 1** can only arise if the two loci are perfectly correlated, which means that only two haplotypes exist in the population: AB and ab

d)

METHOD:

* increase rho
* Varying s (selection) what is the pattern over time
* For one value of S find the median time to fixation of the A allele
* How do the patterns change?

ANS=

(similar to above about the answer about recombination? e)

For recombination the corresponding measures of I D’ I and r2 should have low values indicating that recombination is present.

e)

Q: What kind of selection is this?

ANS:

Positive selection … ??

Techniques: Understanding code and concepts

* Find and highlight through the key words/ parameters and view where they repeat in the whole code
* Repeat for each of the key words
* F1 the important functions in the code
* Understand relationships within the code e.g. how something is defined before and how it is reused
* Step through the code line by line and review the outputs
* Search the key words ‘ FST , coalescent, heterozygosity ‘ LECTURE SLIDES: internet/ google
* Combine past questions into one then research of the answers/ the relevant information
* RESEARCH – example plots in lecture slides; a similar type
* Not more than 3 pages
* Marks available?