Q1)

a)

DO - 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

results as expected?

State assumptions

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

b)

DO –

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

Can check the previous assignment for similarities of differences

expected, why, why not?

c)

where do you think PopX comes from?

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

DO - research Ancestral Recombination Graph (ARG)

e)

Arg as expected, why or why not?

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

All individuals coalesce?

DO - 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

RHO – rate of recombination between the two loci

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)

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)

NO – to.plot = no

Use default values nall = 2000, mu=0

**QUESTIONS**

a)

median fixation A allele?

b)

add recombination (rho)

how change?

c)

new mutation A allele

enters the population and undergoes selection (advantageous for survival)

selected allele A has **initial frequency 0.01**

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

Varying s (selection) what is the pattern over time

For one value of S find the median time to fixation of the A allele

d)

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?

e)

What kind of selection is this?

RESEARCH – example plots in lecture slides; a similar type

Not more than 3 pages

Marks available?

Techniques: Understanding code

* 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