Audrey Bahr
3/13/20

Midterm Part B

1. mult.

W_{11}	W_{12}	W_{22}
1	1.1	1.21

$p_1 = 0.4$ 10 generations

$s = 0.1$

$\Delta p = \left(\frac{s}{1+sp}\right) pq$

Gen	Δp	p
1	$\left(\frac{0.1}{1+0.1(0.4)}\right)(0.4)(0.6) = 0.023$	0.423
2	$\left(\frac{0.1}{1+0.1(0.423)}\right)(0.423)(0.577)$	0.446
3		0.470
4		0.494
5		0.518
6		0.542
7		0.565
8		0.588
9		0.611
10		0.634

freq of A_1 in 10 gen = 0.63

2. mutation s 1 pop: n pop 2: 4n

The mutation is more likely to fix in a small population because it is starting out at a higher frequency relative to what it would be in a larger population.

Pr(fixation) $\approx 1/2N$ Ratio $\frac{1}{8n} : \frac{1}{2n} = 1/4 : 1$

if $(N \cdot s) << 1$

Audrey Bahn
BIO 380P
3/13/20

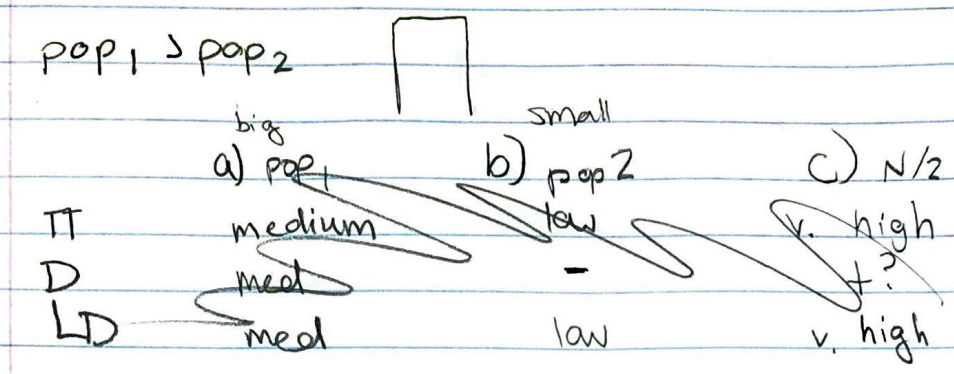
Midterm Exam Part B

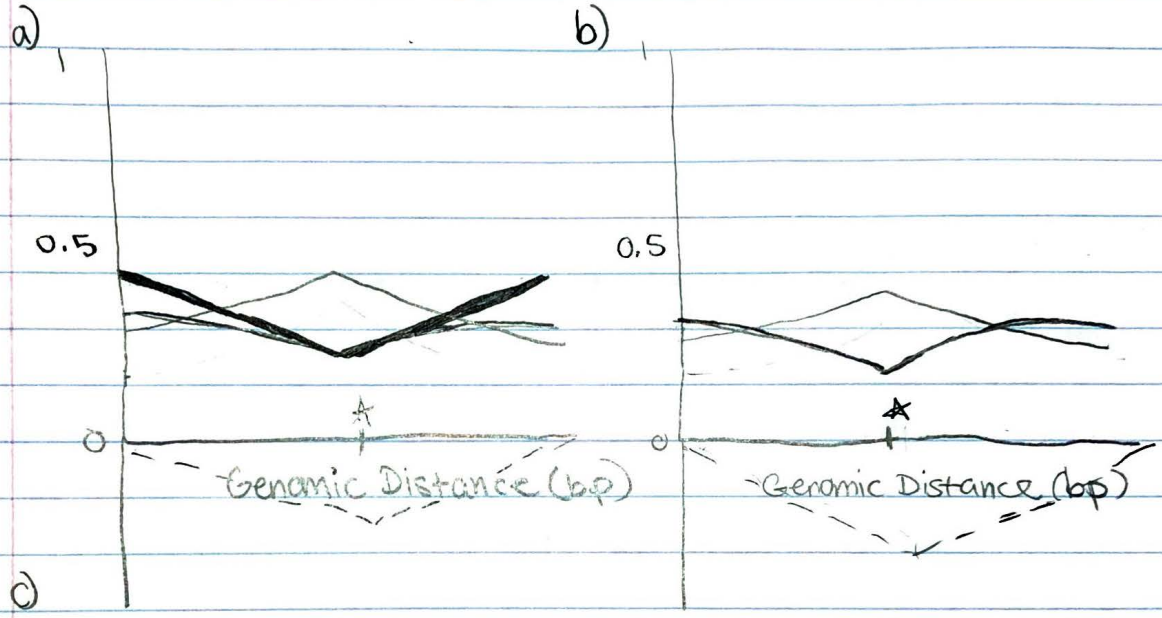
3 Effective Population size: A larger population will have a higher value of π because in a larger population the expected time to coalescence will be longer, allowing more mutations to build up in the gene copies. Population = population & genes

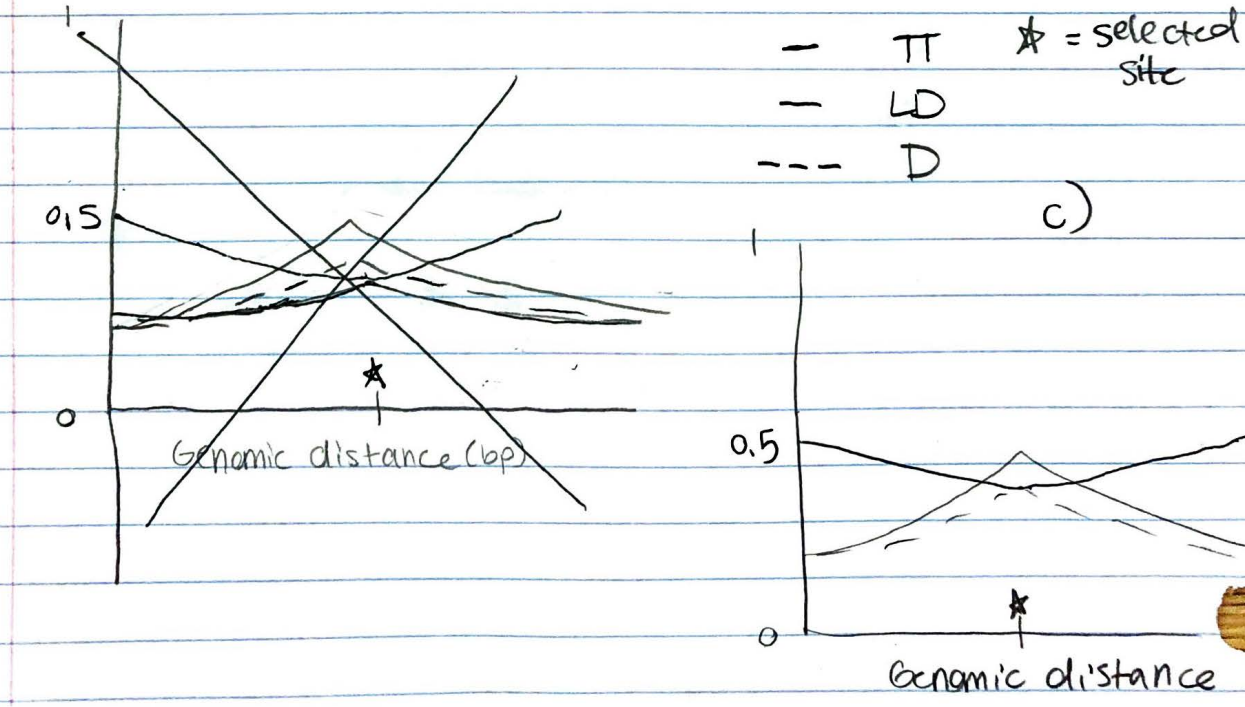
Neutral Mutation rate: A higher mutation rate results in a higher value of π, because more mutations increase the probability that two randomly selected gene copies will differ. Between regions of the genome the rate at which neutral mutations arise in coding regions will be lower because almost all mutations there have deleterious effects, and higher in pseudogenes and intergenic regions which can mutate without having a significant effect on fitness (in general).

Linkage disequilibrium: Regions of the genome with higher LD will have lower genetic diversity. If one locus is selected for, nearby loci can be transmitted along with it through hitchhiking. There will be less recombination in the area, which fails to introduce new combinations of alleles.

4. pop1 > pop2


big
small
a) pop1
b) pop2
c) N/2
TT
D
LD
medium
med
med
low
-
low
v. high
+?
v. high


a)
b)
1
0.5
0
Genomic Distance (bp)
1
0.5
0
Genomic Distance (bp)


c)
1
0.15
0
Genomic distance (bp)
— TT
— LD
--- D
★ = selected site
c)
1
0.5
0
Genomic distance

5

1) You may have forgotten to adjust your p-value for the millions of comparisons, so many sites were significantly associated by chance - Test by lowering the p-value accordingly.

2) Linkage disequilibrium - multiple sites that have no effect on the trait may be linked to the trait that does and show association due to hitchhiking. This could be possible if the trait recently underwent a selective sweep or the associated genes are located on a chromosome inversion, which have low recombination rates. LD can be detected with an integrated haplotype score, measures of correlation, values of π and Tajima's D, and decision-tree methods.

3) Polygenic inheritance - the trait in question, such as height, is contributed to at varias strengths by many SNPs across the genome. This is likely for continuous traits. It can be tested with a polygenic score after LD clumping, and if the PGS can correctly predict the value of the trait in a new sample it is likely to be polygenic.

4) Population structure - If the GWAS data collection fails to take population structure into account (comparing organisms from one area that have the trait vs. those from another area who don't) the sites in the genome may appear to be significantly related to the trait when they are really just associated with the population. This can be tested by taking a more random sample and seeing if most of the associations vanish.