PopGen Midterm

Erik Iverson

3/12/2020

(1) $p = 0.4, \quad q = 0.6$

s for $q = 0.1$

$$q(t) = \frac{q(0)\,e^{st}}{p(0) + q(0)\,e^{st}} \qquad \text{and} \quad e^{0.1(10)} = e$$

Thus

$$q(10) = \frac{0.6e}{0.4 + 0.6e} = \frac{1.6309}{2.0309}$$

$$q(10) = 0.8050 \qquad p(10) = 1 - q(10)$$

$$p(10) = 0.1969 \approx \boxed{0.20}$$

(2) Prob. fix $= U(p) = \dfrac{1 - e^{-4Nsp}}{1 - e^{-4Ns}}$

Given n and $4n$, ratio of probabilities is

$$1 - e^{-4nsp} \; : \; 1 - e^{-16nsp}$$

Erik Iverson
3/12/2020

or $1 - e^{-4x} : 1 - e^{-16x}$

assuming $nsp = x >> 1$,

e^{-16x} is smaller and thus $1 - e^{-16x}$ is larger

Thus, $U(p)$ is greater in the 4n population, fixation probability increases with population size because this reduces the chance of the mutation being lost to drift.

(3) Because $\pi = 4N\mu_n$, the first two factors that affect it are the population size (N) and neutral mutation rate (μ_n). π increases with μ_n ~~because~~ because more mutations, and thus variability, are introduced per chromosome per unit time, and π increases with N because there are more individuals (and chromosomes) on which a mutation can occur. A third factor affecting π is selection via linkage disequilibrium. At loci linked to a locus under selection (same haplotype), π will decrease with

Eric Iverson

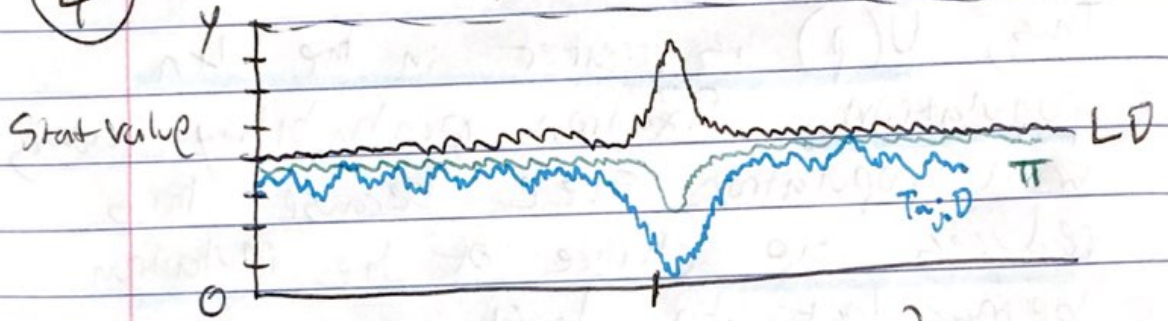
3/12/2020

directional

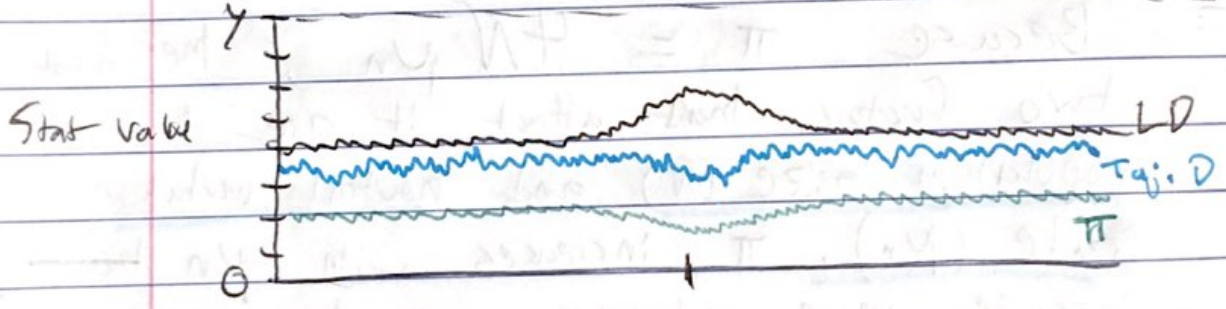
the strength of selection (especially for hard sweeps) as one haplotype increases in frequency and others are lost.

(4)

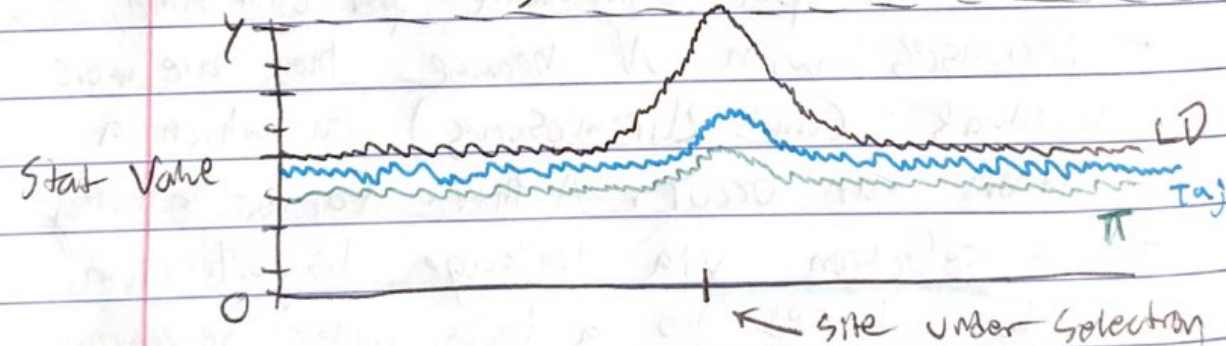
a) Pop 1 (large N_e)



b) Pop 2 (small N_e)



c) Pop 1 + Pop 2



Position in the Genome

Erik Iverson
3/12/2020

(5) Two contrasting situations could be occurring when many or most genes appear significantly associated with a trait in a GWAS- first, the association could be real, and you've discovered a polygenic trait, where perhaps thousands of genes that interact in a network are all associated, to varying degrees (β, effect size) with the trait, some causally and some as downstream byproducts. This is likely if we have an adequate sample size (thousands) for detecting weak associations and we have controlled for population structure. If we have not, we may actually be sampling two or more populations. In this case, the trait may differ between them, but so do thousands of loci, this creates a spurious association between the trait value and these loci, some of which may not even be meaningful or have any fitness effects. To prevent this, use the genomic data to determine the population structure, assign individuals to one, and use population as a covariate.