SELECTION

AARHUS
UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

CALIN PANTEA
PHD STUDENT

TODAY

- Selection – brief lecture recap
- NS textbook exercises – chapters 7 and 8
    - (also in R)
- exercise discussion

FRIDAY PREVIEW

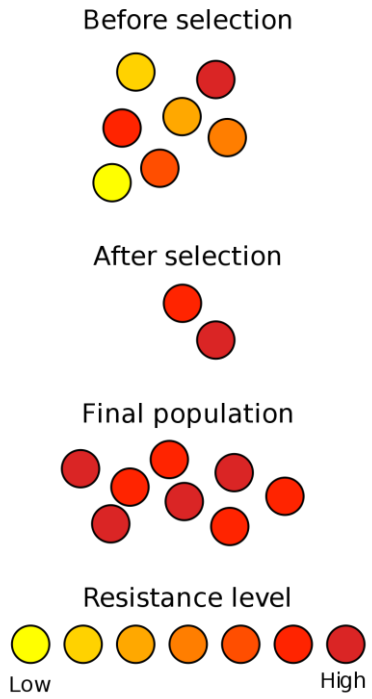
- R exercises – selection (will be up on the course GitHub tomorrow)
- Menti quiz and feedback
- Brief hand-in II discussion

RECAP – OVERVIEW

- viability selection
- types of selection
- selection vs drift
- selection inference: McDonald-Kreitman test and dN/dS ratio

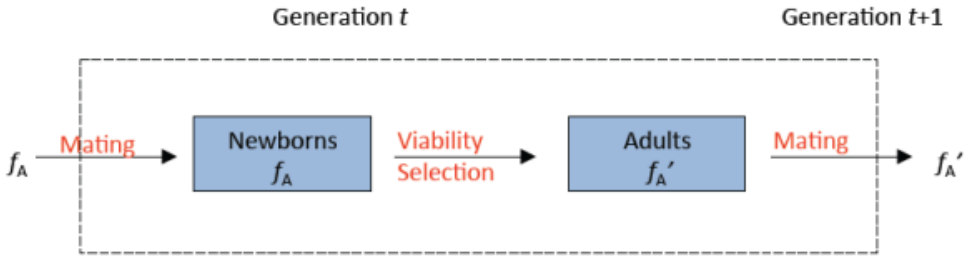
RECAP – SELECTION

- What is selection?
- Why is it interesting?
- How can we quantify it?



RECAP – VIABILITY SELECTION

- selection acting on individual survival to mating



- in diploid organisms:

Genotype	*AA*	*Aa*	*aa*	
Freq @ t	f_A^2	$2 f_A f_a$	f_a^2	
Viability	w_{AA}	w_{Aa}	w_{aa}	
Freq after sel.	$f_A^2\, w_{AA}$	$2 f_A f_a\, w_{Aa}$	$f_a^2\, w_{aa}$	(NOT NORMALIZED)

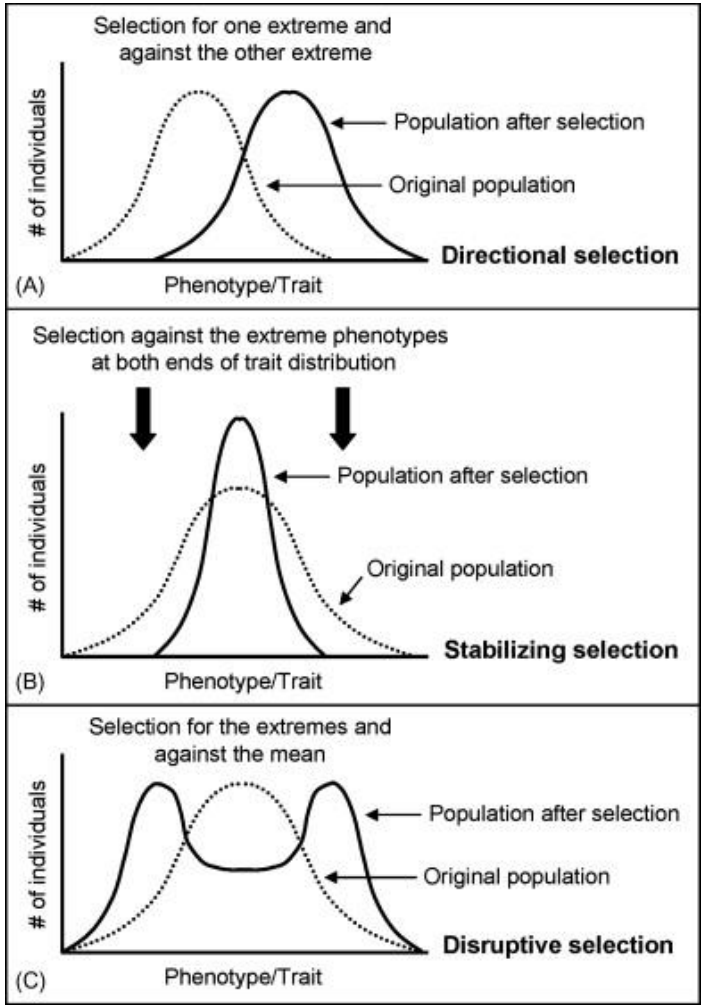
NORMALIZE WITH With $w_{pop} = f_A^2\, w_{AA} + 2 f_A f_a\, w_{Aa} + f_a^2\, w_{aa}$

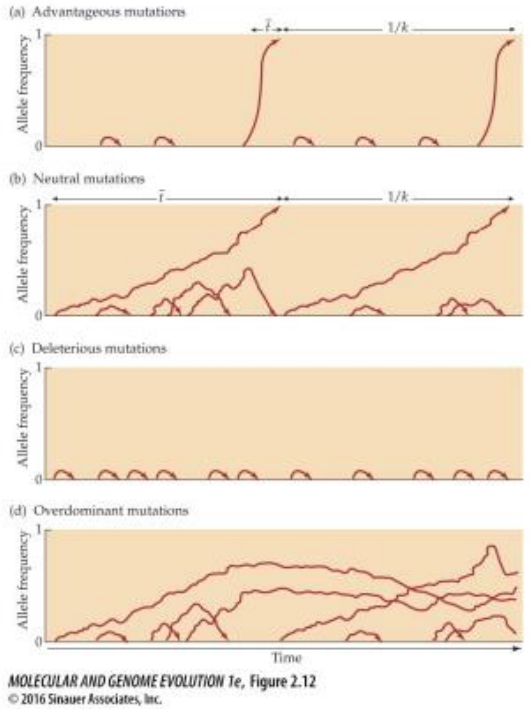
Relative fitness i.e. w_{AA}/w_{pop}, w_{Aa}/w_{pop}, *etc.* allow to predict the change in *allele frequency* due to selection

Rel fitnesses	W_{AA}	W_{Aa}	W_{aa}
	1	1- hs	1- s

RECAP – TYPES OF SELECTION

- Directional selection
  - favours a specific allele
- Stabilising selection
  - favours heterozygotes
- Disruptive selection
  - disfavours heterozygotes



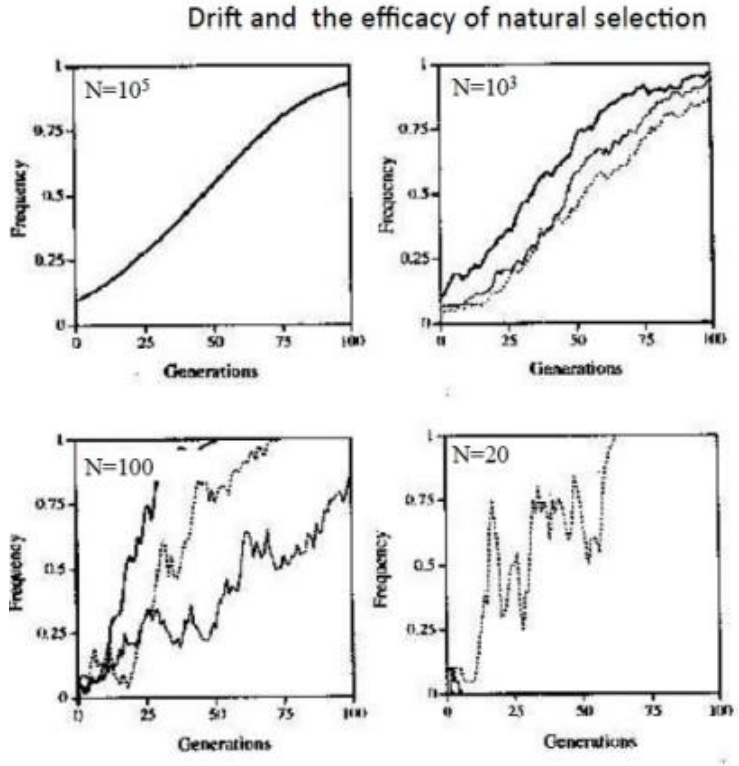


MOLECULAR AND GENOME EVOLUTION 1e, Figure 2.12

RECAP – SELECTION AND DRIFT

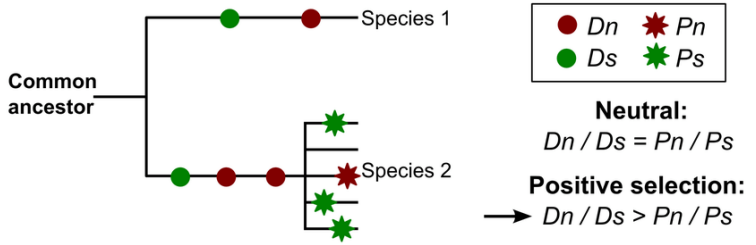
- two independent phenomena affecting allele frequency

- selection acts deterministically and depends on f_A and s
- drift acts stochastically and depends on f_A and N_e



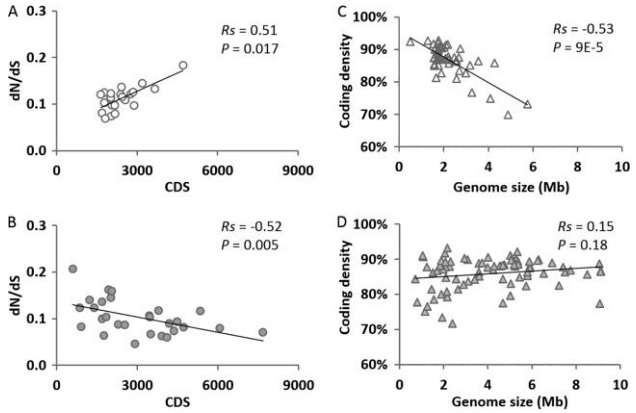
RECAP – SELECTION INFERENCE

- McDonald-Kreitman test (MKT)
  - relies on fixed and polymorphic alleles in coding regions
  - null hypothesis: Dn/Ds = Pn/Ps



	Fixed	Polymorphic
Synonymous	D_s	P_s
Nonsynonymous	D_n	P_n

- D_s: the number of synonymous substitutions per gene
- D_n: the number of non-synonymous substitutions per gene
- P_s: the number of synonymous polymorphisms per gene
- P_n: the number of non-synonymous polymorphisms per gene



- dN/dS (or Ka/Ks) ratio
  - simple ratio of nonsynonymous to synonymous substitutions in coding regions
  - easy interpretation; statistical significance

EXERCISES

Group up and try to solve the exercises from chapters 7 and 8

Until 12:55

Breaktime!

12:55–13:10

EXERCISES

Go through the exercises a little more, then we will discuss them

Until 13:30 (or later if there are few questions)

EXERCISES – CH 7

7.1 Suppose that a new allele *A* is created by mutation in a haploid species and that *A* results in a 1% higher growth rate per unit time.

a. How long will it take for *A* to increase from 10% to 90% in frequency?

b. How long will it take if the growth rate is 0.1% higher?

7.2 Suppose that a mutant allele *A* arose in a haploid population at an unknown time in the past. You know that its frequency today is 0.9. How many generations in the past did the mutation occur if the population size is 10,000? (Ignore the effects of genetic drift.) What about a population size of 100,000?

7.3 Suppose that *A* in Exercise 7.1 has a 5% lower growth rate on average. How long will it take for *A* to decrease from a frequency of 10% to a frequency of 1%?

7.4 Cystic fibrosis (CF) is a Mendelian recessive disease of humans caused by defects in ion transport (OMIM 602421[1]). Until the 1950s, when antibiotics were first used to treat CF patients, most newborns with CF died at an early age. Yet CF is relatively common in Caucasians, with a frequency at birth of 1/2500, which implies that the frequency of CF-causing mutations is about 0.02—a surprisingly high frequency of an allele that is lethal to homozygotes. There is no agreement on the reason for this high frequency.

a. Suppose that an allele that causes CF is maintained by mutation–selection balance. What would be the mutation rate necessary for that allele to have a frequency of 0.02?

b. Suppose that an allele that causes CF is maintained by heterozygote advantage. In order for the equilibrium frequency to be 0.02, what would the difference between the viabilities of the homozygote and the heterozygote have to be?

7.5 The frequency of the A^- allele of *G6PD* in western African populations is about 11%. If A^- is at its equilibrium frequency, what is the selection coefficient against normal (*BB*) homozygotes if the individuals homozygous for A^- have a 50% chance of surviving to adulthood in western Africa? (Ignore the presence of other alleles.)

7.6 A locus with two alleles, *B* and *b*, affects the viability of seeds of a plant population. One-fifth of the *BB* seeds germinate and produce adult plants; 1/6 of the *Bb* seeds germinate and produce adult plants; and 1/10 of the *bb* seeds germinate and produce adult plants. Fertility does not depend on the genotype at the *B*/*b* locus. If the frequency of *B* is ¼ in one generation and the genotypes in that population of seeds are in Hardy–Weinberg equilibrium, what will be the frequency of *B* in the seeds in the next generation?

7.7 *A* is the normal allele at the *β-globin* locus, but in a malarial region of western Africa, the *S* allele of this locus is present at a frequency of 0.2. *SS* individuals have sickle-cell anemia and have only a 10% chance of surviving to reproductive age relative to heterozygous individuals with the *AS* genotype. Normal individuals with the *AA* genotype at this locus have an 85% chance of surviving to reproductive age, relative to *AS* individuals. Assume that at this locus, the genotypes of newborns are in Hardy–Weinberg proportions.

a. If the relative fitness of *AS* individuals is 1, what is the average viability in this population?

b. What are the genotype frequencies among individuals of reproductive age?

7.8 Suppose that a new sand-covered island is created in the Gulf Coast of Florida and that the island is colonized by a population of *Peromyscus polionotus* that is fixed for the dark-colored allele of *MC1R*. If the population grows to 10,000 individuals and then an individual heterozygous for the light-colored allele arrives on the island and mates with one of the residents, how many generations will it take for the light-colored allele to reach a frequency of 99%? (Assume genic selection in favor of the light-colored allele and ignore the effects of genetic drift.)

7.9 Suppose you are concerned with the fertility differences caused by the genotype at a locus with two alleles, *A* and *a*. Suppose that all mating pairs produce the same number of offspring except for the *aa* × *aa* matings, which produce only half as many as the others. All the genotypes have the same viability.

a. If initially the frequencies of the three genotypes are ¼ *AA*, ½ *Aa*, ¼ *aa*, what will the genotype frequencies be after one generation of random mating?

b. Are the genotype frequencies in the newborns in their Hardy–Weinberg proportions?

7.10 Reciprocal translocations occur when there is an exchange of genetic material between nonhomologous chromosomes. Often reciprocal translations have no effect on phenotype, because there is a full complement of genes. But they reduce fertility of heterozygotes by a factor of ½ because half of the gametes produced are aneuploid. Assume that a population carries a reciprocal translocation that has no effect on viability. Let *A* be the translocation. Assume that initially, $f_{AA} = 0.01$, $f_{Aa} = 0.18$, and $f_{aa} = 0.81$. That is, *A* is initially in HWE with frequency $f_A = 0.1$.

a. What are the fertilities of each possible mating pair?

b. Explain why the effect of a translocation results in disruptive selection.

7.11 Suppose the *R* allele of the Rh system to be recessive in its effects, instead of dominant. That is, only *RR* individuals are Rh^+; *Rr* and *rr* individuals are Rh^-.

a. Fill in a table corresponding to Table 7.1 in the text that lists the fitness loss in all types of families. (Recall that incompatibility occurs when an Rh^- mother carries an Rh^+ fetus.)

b. Would this case also result in disruptive selection?

7.12 Suppose a very large plant population has five *S* alleles in equal frequency, 0.2.

a. What are the genotype frequencies in this population if there is random mating? (Hint: there are no plants homozygous for an *S* allele.)

b. Assume that mutation creates a sixth *S* allele. How many more offspring, on average, will the mutant plant have than any other plant in the population?

7.13 The average viability depends on the allele frequencies:

$$\bar{v} = f_A^2 v_{AA} + 2f_A(1-f_A)v_{Aa} + (1-f_A)^2 v_{aa}$$

Draw graphs of $\bar{v}$ as a function of f_A for these three cases: $v_{AA} = 0.5$, $v_{Aa} = 0.4$, $v_{aa} = 0.3$; $v_{AA} = 0.4$, $v_{Aa} = 0.5$, $v_{aa} = 0.3$; $v_{AA} = 0.5$, $v_{Aa} = 0.3$, $v_{aa} = 0.4$.

EXERCISES – CH 8

8.1 Suppose that the mutation rate for single nucleotide change is 2.2×10^{-9} per site per year.

a. What is the rate of substitution of deleterious mutations per million years if the selection coefficient against them is 0.001 in a population containing 10,000, 1000, or 100 individuals?

b. What fraction of the neutral rate are the substitution rates you computed for part a?

8.2 Find the same results as in Exercise 8.1 for an advantageous mutation with a selection coefficient 0.001.

8.3 The fixation probability for a recessive advantageous mutant ($v_{AA} = 1$, $v_{Aa} = v_{aa} = 1 - s$) in a population containing N individuals is $\sqrt{2s/(N\pi)}$. Notice that this probability depends on N no matter how large s is. The fixation probability of a strongly advantageous allele with an additive effect of 0.1% on viability is approximately $2s = 0.002$. How large would s have to be for a recessive advantageous allele to have the same fixation probability in a population of 10,000 individuals?

8.4 Suppose you compare 1000 codons of aligned sequence in humans and chimpanzees. You do not have a computer available, so you examine the second codon positions by hand and find four that differ. What is your estimate of the nonsynonymous substitution rate?

8.5 For insulin, the rate of nonsynonymous substitutions is 0.13×10^{-9} per site per year, and for histones, it is about 10^{-13} per site per year. What is the minimum fraction of the nonsynonymous deleterious mutations that are deleterious if the neutral rate is 2.2×10^{-9} per site per year?

8.6 The rate of synonymous substitution for *β-globin* is 0.8×10^{-9}. In the text, we showed that if a fraction α of the nonsynonymous mutations is strongly deleterious and the rest are neutral, $\alpha = 0.64$. Suppose instead that a fraction α of the nonsynonymous mutations are deleterious but that only ⅔ of those are strongly deleterious. The remaining ⅓ are slightly deleterious with a selection coefficient of 0.001. Use the results from Exercise 8.1 to find α. Assume that the population size is 100 and that the mutation rate is 2.2×10^{-9} per site per year.

8.7 Assume that for *β-globin*, 0.08% of the nonsynonymous mutations are strongly advantageous with a selection coefficient of 0.01 in a population of size 10,000, and that a fraction α are strongly deleterious. If the mutation rate is 2.2×10^{-9} per site per year, what is α?

8.8 Idiopathic torsion dystonia (ITD) is a movement disorder caused by dominant alleles at the *ITY* locus on chromosome 9 in humans. In the Ashkenazi Jewish population, most cases of ITD are caused by a single mutation. On 54 chromosomes with this mutation, 47 have a particular allele (12) at a microsatellite locus (*ASS*) that is 2.3 cM away from *ITY*. The background frequency of allele 12 at *ASS* is 0.086. How many generations in the past did the causative mutation arise? (Use Equation B8.1 in Box 8.5.)

8.9 Disastrophic dysplasia (DTD) is a disorder that causes short stature and unusual growth of the joints in humans. Many cases are caused by dominant alleles at a locus on chromosome 5. In a study of 146 chromosomes from the population of Finland, 139 were found to have an allele at the *CSF1R* locus that is in 3% frequency on non-DTD chromosomes. Estimate the recombination rate between *CSF1R* and the locus that causes DTD under the assumption that the Finnish population was founded 100 generations ago and that one of the founders carried the mutation causing DTD. (Use Equation B8.1 in Box 8.4.) This method for estimating the recombination distance is called linkage disequilibrium mapping.

8.10 Suppose that ten alleles at a gametic self-incompatability locus are present in a plant population in equal frequencies. The model is equivalent to an island model of population structure with ten islands.

a. What is the effective migration rate in this model? (Remember, there are no homozygotes at the S locus.)

b. Use the theory in Chapter 4 for an island population with d demes to predict the expected heterozygosity at a neutral locus at a recombination distance c from the S locus. You have to calculate H_T by averaging the probability that two copies of the neutral locus are on the same S allele background and on different backgrounds.

c. Would you expect the heterozygosity of the neutral locus to increase or decrease if the number of S alleles were larger than 10?

EXERCISES – DISCUSSION

NEXT TIME

- R selection exercises

- Menti quiz for this week
  + feedback

Mentimeter

- Brief hand-in II presentation

AARHUS
UNIVERSITY