6.891 Fall, 2004 Prof. Robert C. Berwick

Outline of Lectures 1-2, 9/8-9/13

1. What does a theory of evolution have to explain about the world around us – if anything?
   - *Adaptation*. Why do organisms 'fit' their environment? (Like a lock and key). This characteristic of 'good design' was a key element of the 'natural theologians' account of the 'goodness' of the Creator's universe. Example: orchid and insect with long tongue than can fertilize it. Darwin's book: *Various Contrivances by which Orchids and Fertilized by Insects,* documented that the intimate relationship between insects and the flowers that they pollinated was matched by their structural compatibility and the insect's behavior. On the strength of his observations, Darwin made bold to predict that the insect that pollinated an orchid (*Angraecum arachnites*, found in Madagascar) with nectaries that were 11 inches long would be a moth with a proboscis that long! 125 years later the moth (*Panogena lingens*) was discovered.
   - *Historical change*. Organisms change over time. (How do we know this? Question: *why* do they change over time?) Is 'evolution' to be equated with 'historical change'?
   - *Complexity*. Why isn't the universe filled with a uniform grey goo? Why do organisms have structure at all – is this an intrinsic property of organisms or an extrinsic one (or both)? Is it just because the environment is complex?
   - *Variation*. Organisms differ. Why are there so many different kinds of organisms? In detail: why do all fish have vertical fins, but all cetaceans (whales, dolphins,…) horizontal fins? (both seem to work equally well). We seem to have two amino acid differences in a certain protein, FOXP2, implicated in motor-neural development, as compared to our nearest relatives, chimpanzees (*Pan paniscus*), and it's been asserted that this difference plays a role in why we talk and chimps don't. Was there 'selection' for the gene that codes for this protein? How could we figure this out? Can we figure this out?
   - *Similarity*. The flip side of variation – organisms resemble each other, at almost any level of description you can name – from the bones in a bird's wing and a person's arm, to the individual nucleotides in their DNA. The *three taxon hypothesis* says that all life can be organized into a 'family resemblance tree' such that any two organisms will have a (least) common ancestor (LCA), and that ancestor will, in turn, be related to the third organism – the LCA and the other organism will themselves have a joint, least common ancestor. This implies evolution by descent – Darwin's theme. Is it true? How can we tell? Much of current genomic research is, of course, predicated on this assumption.
   - *Clumping*. Putting variation and similarity together: organisms seemed to be 'clumped' into groups, no matter how you want to measure them (by their morphology, gene frequencies, whatever). We may imagine drawing a (huge) space with 'notional' axes representing each possible dimension of variation, for example, the set of all genome sequences. No matter what measure we pick, the space will be very, very sparsely filled (do you believe this?).

2. We might depict *any* model for 'evolutionary' change as a *sequence* of states over time, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, … where there is some map, *T,* that carries the state $\boldsymbol{x}$ into a next state $\boldsymbol{x}'$. In general, we say that a *discrete dynamical system* consists of a map *T* from a space $\boldsymbol{x}$ onto itself. (We haven't said anything much yet about constraints that characterize *T* or $\boldsymbol{x}$, e.g., $\boldsymbol{x}$ could be the set of all the genes in an organism, or their frequencies in the overall population of organisms; or it could be a list of all the traits in a (set of organisms) – whatever a 'trait' is - and *T* could be differentiable). As is usual, we may imagine that this map can be iterated: $\mathbf{x}$, $\boldsymbol{T}\mathbf{x}$, $\boldsymbol{T}(\boldsymbol{T}\mathbf{x})$,...$\boldsymbol{T}^{\mathbf{k}}\mathbf{x}$. This sequence is called the *orbit* of $\boldsymbol{x}$. It will be convenient to write the map

in the form, $\boldsymbol{x'} = \boldsymbol{T}\mathbf{x}$, or as $\mathbf{x'-x} = \boldsymbol{T}\mathbf{x-x}$. Equations of the second type are called *reccurrence* or *difference equations.* We give an explicit example shortly.

3. The general problem is one of constructing a state space that will be dynamically sufficient, and a set of laws of transformation in that state space that will transform all the state variables. The laws also have to be empirically sufficient. This entails an interchange between finding the laws and picking the right state variables, to predict a future state $\boldsymbol{x'}$ given a current $\boldsymbol{x}$. The transformational laws cannot be arbitrary – usually they contain some parameters P that are not themselves a function of time or the state of the system. Second, the laws contain the elapsed time except in the case of equilibrium, and may or may not refer specifically to the absolute time t depending on whether the system carries in its present state some history of the past. Finally, most importantly, the laws of transformation must contain the present state of the system and suffice to produce the next state. Example – you can't predict the future position of a satellite from its current position alone. You need to know 3-D position, velocity, and acceleration – 9 variables in all, and that is dynamically sufficient.

4. So what about for evolutionary theory? What is the state space? What are the laws of transformation? It is important to stress that we can't go out and describe the world any way we want and demand that an explanatory and descriptive theory be built on that description – it might be dynamically insufficient. This has important consequences, because to model evolution we will necessarily build such dynamical systems – but this takes some care. What kind of 'force' is evolution? What are the state variables?

5. Much of the remainder of this course will consist in characterizing and exploring the dynamical system behavior of T and $\boldsymbol{x}$**:** rate of evolution; convergence times; stability properties; measurement and inference problems – how can we tell from a current state, what has happened in the past; how can we measure the 'forces' of evolution; how various characterizations of T and $\boldsymbol{x}$; how sensitive our these models to our assumptions.

6. We may identify two distinct 'modes' of accounting for evolutionary change, *transformational* and *variational.* Transformational theories have been the most widespread for explaining historical change. The prototypical transformational example is stellar evolution. In transformational evolution, the properties of each individual in a group change, and, as a result, the overall group compositional properties change. This explanatory mode was also common to perhaps the most popular view of biological evolutionary change in the 1700s, given the name *Lamarckism* (after Jean-Baptiste Monet, Chevalier de Lamarck), although he did not originate this approach. Like stellar evolution, on this view biological change too is transformational: all individual organisms making up a population change – each giraffe's neck gets longer reaching for the treetops – and as a result of *individual* change, the population property in question as a whole changes.

7. In contrast, Darwin seems to have been the first to introduce a completely novel account for evolution, the *variational* mode of evolutionary explanation. In a variational story, no single *individual* undergoes evolutionary change within its lifetime. Rather, *already existing* variation among individuals is *selected* (sieved) and the group population property as an ensemble changes because different individuals are selected to make up the group. Or: In the variational mode, there is variation among the individual units comprising the whole system. The system changes in time by a change in the proportions of the different kinds of units, as a consequence of differential survival and reproduction of the units. In Darwin's scheme, the group as a whole changes through time because objects with different properties leave different numbers of descendants.

8. It is important to understand that different domains of phenomena evolve by different modes. The transformational mode, which is correct for galaxies and embryos, is not correct for species.

9. The essential nature of the Darwinian revolution was not the introduction of evolution as a worldview. This is historically not the case. It is rather the replacement of a metaphysical view of variation among organisms by a materialistic view. Darwin's materialistic view replaced the Aristotelian, Platonic conception that there is a single ideal 'type' – an ideal squirrel, or whale, or E. coli – and so, and that the 'failings' of organisms in their perfection of adaptation somehow to be related to their failure attain this ideal – they are imperfect approximations to their 'true nature.' The failure of individual cases to match the ideal was a measure of the imperfection of nature. (Compare this to the Newtonian idealizations about frictionless planes, etc.) Darwin overthrew this metaphysical picture and replaced it with a materialistic one: what matters in Darwinian evolution by natural selection is the actual variation in actual individuals. This individual variation is of the essence – as we'll see many times in the course: the *rate* of evolution by natural selection is *directly proportional* to the amount of standing variation. Variation is the fuel that evolution burns. (And in addition, unlike a fire, evolution has to 'make' its own fuel.)

10. So: Darwin's explanation for the 'clumping' that we see in the biological world is the conversion of the standing variation between individuals into variation between species and groups.

11. All 'vulgar' versions of evolution at the time of Darwin thus already included variation and inheritance. Darwin added the key notion of *selection*, dubbed *natural selection* by analogy with the *artificial* selection that Darwin observed amongst pigeon, dog, horse breeders. Selection 'sieves' existing variation by letting 'pass through' individuals with certain properties, but not other properties – like panning for gold. Inheritance then transmits these individuals, with their selected properties, to the next generation. These comprise the three key components of Darwin's theory:
    a. *Variation*: Among individual members within any population, there is variation (in genes, morphology, physiology, behavior, …)
    b. *Heredity* (information transmission): Offspring resemble their parents more than they resemble individuals to which they are unrelated
    c. *Selection*: Some forms are more successful at surviving and reproducing than other forms in a given environment. (aka, the 'misnomer' "survival of the fittest")

    It is a valuable exercise to consider whether these three components are necessary and sufficient for evolution by natural selection – alternatively, whether it is possible to design some *other* system that would work to create the living world around us. At least superficially, it should be clear that without variation, evolution can do nothing – there is nothing that selection can do to separate out a group with different ensemble properties. Similarly, without some form of inheritance, a differentially selected set of individuals would not be 'passed on' to future generations. And of course, some kind of selection must be part of 'natural selection'.

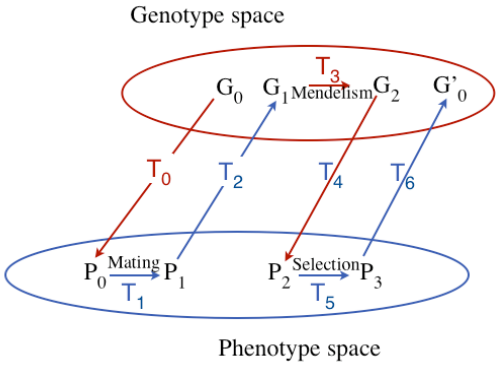
12. But what *sort* of variation, heredity, and selection? Consider heredity. There are interesting constraints on the kind of 'information transmission' systems that can (easily) lead to evolution by natural selection. One of these constraints was first driven home in perhaps the most devastating review of *Origin of Species* – at least the one that troubled Darwin the most - by the Scottish engineer Fleeming Jenkin. This EECS pioneer almost sunk Darwin's theory. Jenkin's point had to do with the nature of inheritance. In Darwin's time, the most widely accepted notion of 'inheritance' was essentially that of *blending inheritance* - 'mixing by

blood' – like blending paint (a very old idea). But blending inheritance runs afoul of natural selection's demand for variation. It is easy to see that if in fact inheritance worked by blending, then any variation would be quickly swamped within a few generations – in fact it is halved at ever generation – and so turned into a kind of uniform brown mud. We wind up with nothing for natural selection to 'select'. Darwin himself was unable to refute this argument, and wound up embracing Lamarck's inheritance of acquired characteristics, along with a theory of 'hyper mutation' that would pump in enough variation to keep selection going (if ½ of all variation is lost each generation, then ½ the variation we see in the current generation must be due to mutations introduced by the immediately preceding generation).

13. The problem was resolved by Mendel's discovery that inheritance is quantal, or particulate in nature, and not blending (the particles of course ultimately turned out to be genes). R.A. Fisher (1911, 1930) demonstrated exactly how blending inheritance would halve variance at each generation: Let x denote the deviation of one parent from the mean of any trait (e.g., height, or "Paul Newman blue eyes"), and y denote the deviation from the mean of that trait for the other parent. The variance of the trait is the expected (mean) value of the square of deviations from the mean, or $E[x^2]$. On the model of perfect blending inheritance, then the deviation of the offspring would be ½(x+y) – exactly half-way in between. We can now calculate the expected value of the square of this deviation: $E[\{½(x+y)\}^2]$=E[¼ $(x^2+2xy+y^2)$]. The quantity E(xy) must be zero, since the expected value of deviations from the mean is always 0. The expected value of x^2= the expected value of y^2, so we have, E[¼ $(x^2+2xy+y^2)$]= E[¼$(2x^2]$=E[½ (x^2)]= ½ E[x^2] – that is, exactly half the original variance. Thus under blending inheritance variance declines exponentially with each generation. We also must show how 'particulate' inheritance removes this problem. Intuitively, if inheritance works via 'quanta' that we call genes, and these genes are not blended – neither created nor destroyed – during reproduction, then the variance remains constant (all other things like mutation being equal) – we can juggle a handful of different colored jelly beans around, and mix them, but the individual numbers of different colored jelly beans will remain the same. This can be demonstrated more precisely via the so-called "Hardy-Weinberg laws", which we turn to below. Sadly, though Mendel's work was published during Darwin's lifetime – Darwin even had a copy of Mendel's publications – it seems that he never read about them, and the articles were found uncut (unopened) in Darwin's library at the time of his death. Mendel's algebraic system evidently did not appeal to Darwin, who claimed in his autobiography to be all in a muddle about mathematics.

14. We next pursue an elaboration of the notion of a dynamical system mapping as our model of evolutionary change. We describe any (array of) organisms via two key descriptive spaces: Genotype space and Phenotype space. By "genotype" we mean the full array of genes in an organism or set of organisms (for now, think of it as a very long vector, say, 20,000 elements in the case of humans.) For now we don't say what the 'axes' are in this space. Genes occur in variants, called *alleles*. The variants within a species, over all genes, specify distinct *genotypes*. For example, we saw that there is a particular gene that specifies normal human hemoglobin in red blood cells, and another, which differs in exactly one DNA 'letter' (the nucleotide adenine, A, is replaced by a thymine, T), whose corresponding protein has a valine instead of a glutamic acid, which in turn yields a hemoglobin that forms 'bent' crystals and causes red blood cells to become crooked and 'sickle' shaped instead normally elliptical. These are two *alleles* of the gene for hemoglobin (more precisely, the beta-chain of one part of hemoglobin), or two 'genotypes'. By "phenotype" we mean the 'form that shows' – i.e., the actual biological form that an organism presents to the world (and so casually interacts with it in the sense of affecting the outcome of selection). In our example, the two distinct hemoglobins (and consequently distinct red blood cell shapes) are two different 'phenotypes'. Of course, genotypes can also differ between species as well as between individuals. A genotype, then, is simple the string of nucleic acids, the 'genetic code' that make up the DNA

of an individual organism. Selection acts on phenotypes (note that we have been a bit vague as to whether a phenotype might include the notion of a genotype or DNA sequence – doesn't this casually interact with the world? More later.) In the case of normal vs. sickle-shaped red blood cells, we know that the latter are extremely debilitating and lead to early death – the sickle-cell phenotype is 'selected against'. We stress that a full evolutionary theory must pass back and forth between these spaces by means of (as yet) unspecified, and current unknown, mappings. Here is a picture:



15. If we start off in genotype space, in some state G_0, then we must pass back and forth between genotype space and phenotype space several times in order to reach the next possible genotype state G_0': first, the genotype must be realized as a phenotype P_0; (the actual organisms) – this via a set of 'developmental laws' T_0 that turn genotypes into organisms; then these organisms must meet and mate, again via some set of laws T_1; then the mated pairs produce the 'raw' DNA, the *gametes* (e.g., egg and sperm), which brings us back to genotype space via another mapping T_2 specifying how gametes get produced from mated organisms; the genes then reassort and combine via the rules commonly known as Mendelism, T_3; then the fertilized egg(s) (a *zygote*) develops into offspring, via developmental transformations T_4; natural selection then acts on the offspring, winnowing them, T_5; then finally these surviving offspring produce an array of genotypes (via their games, i.e., sperm and eggs) which become the new G_0' to start the process all over again. Whew. It almost goes without saying that except in extremely limited cases we have no knowledge whatsoever about these T's, or how to compute them. We can say, however, that our theory will have to be both dynamically and empirically sufficient. As it stands, to do this would result in a theory concerning a vastly more complex domain than any yet dealt with by physics or molecular biology.

(Aside 1: We should remark that even under naïve assumptions the size of these spaces might be enormous – for example, suppose we take genotype space to be 'discrete' and consist of the set of possible gene sequences for an organism. If there are 10,000 genes in an organism's gene sequence – not far off - with 3 gene types or 'alleles' each, that amounts to $3^{10,000}$ possible genotypes or 'points' in this space.) (Aside 2: as correctly pointed out, we really ought to included *another* space in this analysis, namely, the external environment or context in which G and P reside.)

16. If we have no knowledge of these *T* mappings then how can evolutionary biology proceed? In fact, there are two moves that are made, the obvious ones. We can either 'assume away' P, and do our modeling only in genotype space; or we can assume away G, and do our modeling only in phenotype space. Both strategies are adopted. We can collapse G=P, and work entirely in Phenotype space. This is the province of biometrics – we put this to one side for now. Or, we can collapse G and P, and work entirely in genotype space. This is the strategy of most of evolutionary population genetics: to study the origin and dynamics of genetic variation within populations.

17. While this is a much more modest goal that all of evolutionary theory, it is an essential ingredient. If we adopt this subgoal, we are in effect equating evolution as "change in gene frequencies", as is common in this approach. However, while population genetics has much to say about changes or the stability of the frequencies of genes in populations and about the rate of divergences in gene frequencies in populations, it has contributed little to our understanding of speciation and nothing to our understanding of extinction.

18. Still, the sufficient set of state variables for describing an evolutionary process within a population must include some information about the statistical distribution of gene frequencies. It is for this reason that the empirical and mathematical study of population genetics has always begun with and centered on the characterization of the genetic variation in populations.

19. We begin then by adopting the "evolution as change in gene frequencies" view, and describing the simplest dynamical system modeling this. We want to know what theory says about the reproduction of genotypes in a population. This results in the derivation of the so-called *Hardy-Weinberg proportions*. We imagine a population reproducing without any natural selection or any interference by any other forces such as mutation or migration. The Hardy-Weinberg result serves as a kind of "Newton's First Law" for the genetics of evolving populations, because it says that under such conditions gene frequencies (and their variance) will "remain at rest" – that is, gene frequency proportions will *remain in equilibrium* as long as there is no other force to disturb them. We use this as a 'baseline' model and then introduce selection, migration, etc. as 'forces' that displace a population from its equilibrium. H-W is the second half of the demonstration that Mendelism actually goes hand-in-hand with Darwin's theory – it is virtually a necessary part of Darwinism, since it serves to maintain variation unless there is some other force to disturb it. It would be an interesting exercise to see whether one could develop an alternative that could replace Mendelism, and still get the conditions for the evolution of complex life.

20. The Hardy-Weinberg "law" is based on following assumptions:
    - A single random mating population
    - Infinitely many individuals (Why do we need this assumption? Follow it out in the analysis below)
    - No mutation
    - No selection (no differential fertility, viability)
    - No immigration or emigration
    - Non-overlapping generations

21. Suppose we have one gene that comes in two variants, or alleles, denoted *A* and *a*. If we have genotypes with current genotype frequencies *P*, *Q*, and *R* of genotypes *AA, Aa*, and *aa*, they have a fraction $p = P + 1/2\ Q$ of their genes being *A* rather than *a*. The value *p* is the gene frequency (note the difference between gene frequency and a genotype frequency, e.g., *P*). The gene frequency of the *a* allele is, for the same reasons, $q = 1/2\ Q+R$. These can also be computed by counting the fractions of *A* and *a* among the individuals.

22. Random mating is equivalent to random union of gametes. Imagine making a pot of female gametes, a pot of male gametes, and drawing a pair, one from each. The equivalence comes because a random member of the offspring generation is descended from a random female and a random male, and Mendelian inheritance ensures that the gametes each contributes contain a random one of the two copies (at this locus) in that individual. Drawing a random parent, and then having it choose one of the two copies by Mendelian segregation, is equivalent to drawing one of the copies from the population at random. [Indeed, the *variance*

of this draw, in the case of just two alleles in fractions p and $(1-p)$ is $\frac{1}{2}\, p(1-p)$ as per standard sampling theory from a binomial distribution, a result we shall draw on below – or refer to appendix A in your Sean Rice textbook.

23. The probability that the offspring gets an *A* from the female parent is p, and the probability that it gets an *A* from the male parent is also p. Because these are independent as a result of random mating, the probability that the individual is *AA* is then p^2.

24. The result is that AA, Aa, and aa have expected genotype frequencies p^2, $2pq$ and q^2.

25. The gene frequency in this offspring population is again p, since in that generation P is p^2 and Q is $2pq$, so that $p = P + 1/2\ Q = p^2 + 1/2\ (2pq) = p$.

26. If we again mate these individuals randomly, the gene frequencies in the second generation are again p and q.

27. Thus the genotype frequencies become these "Hardy-Weinberg" proportions, and stay that way forever. The gene frequencies remain forever p. If we are talking about just *one* gene 'location' (= "locus") then the frequencies for two alleles are forever p and $(1-p)$.

28. Mendelian genetic systems thus do not tend to lose genetic variability just because of random mating. Blending inheritance would lose it. The fundamental reason is that segregation in a heterozygote yields gametes that are 1/2 *A* and 1/2 *a*, whereas in blending inheritance it is as if they were all medium-sized *A*'s. [Q: what would happen in a non-Mendelian world? Variation lost? No evolution of complex life?]

29. When we relax the assumption of no differential viability and no differential fertility, we now have natural selection going on. We can count genes to *define* the notion of absolute fitness.

30. Let us now consider the *simplest* case, with just *one* gene (one 'locus') that can take one of two possible forms (alleles); we generalize this immediately to the case of many alleles (generalizing to multiple loci is more complex). We *retain* all the other assumptions about infinite population size, etc. It is *extremely* important to think about the biological reality of these assumptions, and what effect they have. Note that we are, in effect, assuming that the genotype-phenotype mapping is *direct* (i.e., all those T's are identity functions). We also assume that a single gene's contribution to the outcome of selection may be calculated, or 'factored out' no matter what its interactions with other genes. That is, the effect of the gene appears directly in the phenotype, and thus selection can directly affect it.

31. We may imagine a string reaching between a gene and its phenotypic realization: if we pluck the string at either the gene or the phenotype end, the corresponding other part wiggles. If there is just a single string, with gene and trait at either end, this is easy: this is the case (roughly) in sickle-cell anemia. But the world is more complex. This is a familiar situation in the AI literature, when we want to construct what are known as 'causal models': suppose a number of factors conspire to produce a particular outcome and we want to draw inferences about 'who is to blame' (the 'credit assignment problem') – e.g., what caused the great Chicago fire? The biological reality is that there are typically many genes that act in concert to produce a given surface trait, and, importantly, vice-versa. Example: you're all familiar with the 'textbook' idea that eye color is produced by 'crossing' B(rown) and b(lue) eye types – which looks like a locus with two alleles. In reality, there are at least 12 different enzymes that yield eye color; the basal mammalian eye color is Paul-Newman blue, which in fact is the *absence* of any color. Of course, it's worse (or better) than that: typically genes and traits are in a many-many relation. What does that say about natural selection 'twiddling' a trait? How

does it reach down and pluck on a gene's string? The genotype-phenotype relation is many-to-many.

32. Please bear in mind all this as we pursue the simplest possible model: genes are *pleiotropic* (one gene can have functions – like a Swiss army knife - e.g., when you do an expression analysis for a gene, it can play many roles: for example, the FOXP2 gene mentioned as implicated in neural development also seems to affect bone growth). Further, genes are *epistatic* – they interact to yield blue or brown eyes.

33. That said, the *absolute fitness* of each genotype is the expected contribution a newborn individual of that genotype makes to the next generation. This is the product of 1/2 (viability)(fertility). The one-half is because each offspring it has only gets one-half of its genes from that parent.

34. Many populations are subject to density-dependent population size regulation. If we can assume that this falls "fairly" on all genotypes, then it simply multiplies all viabilities by the same number, and/or multiplies all fertilities by the same number. It will do this if the density-dependent population size regulation acts at a different life stage, in a way unrelated to whatever causes the other fitness differences.

35. If this is true, then the ratios of the absolute fitnesses do not change as the population changes density, only the multiplier that makes them into absolute fitnesses.

36. Then we can define *relative fitness* of a genotype as the ratio between its absolute fitness and the absolute fitness of some reference genotype. Thus relative fitnesses might be 1 : 0.8 : 0.7 for the three genotypes, for example, when *AA* is the reference genotype. We will use *w* to denote these relative fitnesses, subscripted as required to refer to the genotype in question.

37. With this one-locus (=one gene) case, one can compute the gene frequency after natural selection. The genotype frequencies at the beginning of the generation are of course in the assumed H-W ratios of $p^2 : 2pq : q^2$. When we count them by their contributions to the next generation (as a result of differential survival and fertility) they are in the proportions $p^2 w_{AA} : 2pq w_{Aa} : q^2 w_{aa}$. (Alternatively, we use the notation w_{11}, w_{12}, and w_{w22} for the relative fitness values corresponding to *AA*, *Aa*, and *aa*.

38. These three numbers don't add to one, usually. So we can make them into frequencies by dividing by their sum. The sum is the *mean fitness*, denoted $\bar{w} = p^2 w_{AA} + 2pq w_{Aa} + q^2 w_{aa}$, which is also the average value of the relative fitness of a randomly chosen newborn. [Remark. Note that we slide between using 'frequencies' as a proxy for 'probabilities' – connect this with our assumption of an 'arbitrarily large' population.]

39. Dividing the sum by $\bar{w}$, we get the three frequencies: (NB – we are using frequencies as proxies for probabilities again):

$$\frac{p^2 w_{11}}{\bar{w}} : \frac{2pq w_{12}}{\bar{w}} : \frac{q^2 w_{22}}{\bar{w}}$$

40. Taking the frequencies of these three after selection (i.e. according to their contributions to the next generation) the gene frequency of *A* in that next generation will be the frequency of *AA* plus half that of *Aa*:

$$p' = \frac{A \text{ survivors}}{\text{all survivors}} = \frac{p^2 w_{11} + \frac{1}{2} \times 2p(1-p)w_{12}}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{11}} \text{ or}$$

$$p' = \frac{p(pw_{11} + p(1-p)w_{12})}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{11}} = p\frac{\bar{w}_1}{\bar{w}}$$

$$p' - p = \frac{p(1-p)\{w_{11}p + w_{12}(1-2p) - w_{22}(1-p)\}}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{11}}$$

41. Note the rightmost expression: it says simply that the new gene frequency is the old one (p) times the mean fitness of the genotypes that a randomly-chosen A allele happens to find itself in ($\bar{w}_1$), divided by the mean fitness of everybody. In short, the gene frequency will increase if the mean fitness of A's is greater than the mean fitness of random individuals.

42. We can now compute the *rate of change of gene frequency as a result of natural selection*, i.e. *dp/dt.* We consider the dynamical system equation (the recurrence relation) described by the recurrence

$$p' = \frac{p^2 w_{11} + p(1-p)w_{12}}{p^2 w_{11} + 2p(1-p)w_{12} + (1-p)^2 w_{22}}$$

43. Calculating *p'-p* we get the *fundamental recurrence formula*:

$$p' - p = \frac{p\bar{w}_1}{\bar{w}} - \frac{p\bar{w}}{\bar{w}} \Rightarrow \Delta p = \frac{p(\bar{w}_1 - \bar{w})}{\bar{w}} \text{ with 2 alleles,}$$

$$\bar{w} = p\bar{w}_1 + (1-p)\bar{w}_2, \text{ so substituting:}$$

$$\Delta p = \frac{p(1-p)(\bar{w}_1 - \bar{w}_2)}{\bar{w}} \text{ and now substitute for } (\bar{w}_1 - \bar{w}_2) = \frac{d\bar{w}}{dp} =$$

$$\Delta p = \frac{p(1-p)}{\bar{w}}\frac{d\bar{w}}{dp} = \frac{p(1-p)}{}\frac{d\ln(\bar{w})}{dp}$$

44. Let's take a look in more detail at this last basic equation, first formulated by Sewall Wright (1930). Note that Δp, the change in allele frequency, hence the *evolutionary change,* depends on two components that get multiplied together: (1) the factor $p(1\text{-}p)$, which is the *genetic variance*, or *heterozygosity*; and (2) $\frac{d\bar{w}}{dp}$, which is the *gradient* of the mean fitness with respect to allele frequency p. Thus the first factor gives the *amount* of change, and the second factor says *which direction* it is applied in. The direction of change – whether an allele increases or decreases in frequency - is defined by the equivalent term $\frac{d(\ln \bar{w})}{dp}$ - the slope of the plot of $\ln \bar{w}$ as a function of p. Since $\ln \bar{w}$ is the instantaneous per capital rate of growth of the population, it is as if the population is climbing a slope defined by the population growth rate. We may imagine, as Wright did, that this describes a 'fitness surface' or 'adaptive landscape' and that we move in a direction so as to increase mean fitness – drive it

uphill. The picture is a useful one because it underscores all the strengths and weakness of this 'evolutionary search for maximum fitness:

- The search is local (i.e., best-first)
- The search is not guaranteed to find a global maximum (and indeed could get stuck on a local maximum)
- The search is not guided by some knowledge of the final 'goal' – natural selection is opportunistic
- The search cannot go backwards (that is, mean fitness always increases – see below)

45. How do we get this interpretation? The $p(1\text{-}p)$ factor comes from binomial sampling theory: we are drawing one allele from a very large population N, so the variance in the probability of getting frequency p is $p(1\text{-}p)$. This variance is multiplied by the "force" of selection – the gradient in fitness. In other words, this variance is changed by the derivative of a potential function, $\frac{d(\ln \bar{w})}{dp}$. Alternatively: it is *multiplied by the slope* of the fitness function and divided by the mean population fitness.

46. Without saying anything else, we can already tell that the change in allele frequency will be large at intermediate values where the variance term is large (the maximum rate of change where p=q=0.5), and ever smaller near allele fixation (i.e., as p approaches 0 or 1).

47. We can thus (approximately!) write down a formula for the time to move from one allele frequency to another:

$$dp = kp(1-p)dt \text{ or}$$

$$\int \frac{1}{p(1-p)} dp = dt \text{ so}$$

$$t(p_1, p_2) = \int_{p_1}^{p_2} \frac{1}{p(1-p)} dp$$

48. We can glean a lot just from this form. This tells us that in general it will be very difficult to catch natural selection operating in *flagrante delecto,* as it were. As the frequency of p approaches unity, the time required for even small changes will be small, owing to the term $1\text{-}p$ in the denominator. We also note that the equation has the form of a logistic (we can integrate by partial fractions). (Aside: note importantly that we are also assuming that the relative fitnesses remain constant! We examine deviations from this assumption below.)

49. Further, we know that mean fitness will always increase: i.e., $\Delta\bar{w} \geq 0$ - just subtract two w's, we get:

$$\Delta\bar{w} = 2p(1-p)\{w_{11}p + w_{12}(1-2p) - w_{22}(1-p)\}^2$$
$$\times \{w_{11}p^2 + (w_{12} + \frac{1}{2}w_{11} + \frac{1}{2}w_{22})p(1-p) + w_{22}(1-p)^2\}\bar{w}^{-2}$$

which is non-negative. So, on this model, natural selection acts so as to increase or at least maintain mean fitness, at least in accord with our intuitions.

50. Let's see if we can further understand the dynamics of this recurrence equation – where are the equillibria points, given various values for w and p; how much time (number of

generations) does it take to reach an equilibrium, etc.? First, it is clear that there are fixed points at p=0 or 1. We can understand matters by comparing the fitnesses of *AA, Aa,* and *aa*.

51. If there is just one locus (gene), here it is helpful to plot the "adaptive landscape," $\bar{w}$ against *p,* recalling the equation:

$$\bar{w} = p^2[(w_{11} - w_{12}) + (w_{22} - w_{12})] - 2p(w_{11} - w_{12}) + w_{22}$$

52. This is the equation of a parabola, and its slope, $\frac{d\bar{w}}{dp}$ evolution's 'guiding hand'. We can examine the possible fixed points and interior equillibria conditions via cases. There turn out to be seven distinct possibilities (ways of drawing this curve) – three lines and two parabolas.
Cases (1)–(3): If w_{11}=w_{12} and w_{22}=w_{12}, then the p^2 factor is 0. In this (degenerate) case, $w_{12} = \frac{w_{11} + w_{22}}{2}$, and $\bar{w}$ is linear. (See Populus try it out). If all three fitnesses are equal, selection does not operate, $\bar{w}$ is constant, and all points are equillibria. Otherwise, the slope of $\bar{w}$ is nonzero, and p converges to 0 or 1. Thus, the homozygote with the highest fitness gets established. [this makes 3 different linear graphs in all].

Cases (4)-(7): In the generic case, $w_{12} \neq \frac{w_{11} + w_{22}}{2}$, $\bar{w}$ is a parabola whose extremum is at the point:

$$\hat{p} = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})}$$

Here there are four subcases. If the fitness of the heterozygote w_{12} is between the two homozygotes, w_{11} and w_{22}, then

$$(w_{11} - w_{12})(w_{22} - w_{12}) \leq 0$$

Subcase 1: Suppose *AA* is dominant, w_{12} =w_{11}. Then $\hat{p}$ = (w_{22}-w_{12})/(w_{22}-w_{12}) = 1. (*A* increases until it fixes at 1.0)
Subcase 2: Conversely, when w_{12} =w_{22}, the numerator is 0, and $\hat{p}$=0. (See Populus graphs 4 and 5). It's clear from the way the gradient points – which way the wind blows.

Subcases 3 &4: If the heterozygote fitness is *not* between the two homozygotes, the equilibrium is in (0, 1). There are two possibilities: (i) the heterozygote fitness is greater than both homozygotes, *or heterosis*, AKA *overdominance,* AKA *heterozygote superiority*: w_{12}> w_{11} and w_{12} > w_{22}. Then we have figure 7 – no matter where we start, we get to a maximum somewhere in the interior. This corresponds to a (simple version of) the sickle-cell case in malarial regions. (But not quite, as we shall see…) (ii) w_{12}< w_{11} and w_{12} < w_{22} i.e., *underdominance* – the heterozygote fitness is less than both homozygous fitnesses. Then the fixed point depends on which side of the minima we start at (we can show that the mapping *T* is monotonic – take the derivative and it's positive). Note that case (i) is the only one where the heterozygote is maintained. Is this why we see so much variation? It was thought so at one time…but this turns out not to be true.

53. What of the relation between *p* and *t* (time *t* measured in generations)? We can carry out the recurrence calculation directly, and for the cases of dominance (*w* for *AA* and *Aa* a factor *s* greater than that for *aa*) and no dominance (all *w*'s equal), to obtain the following table).

54. Table for fixation: # generations spent in various frequency ranges of *p*: [Question: why does it take approximately 200 times *longer* to eliminate *a* in the case where *A* is dominant?]

Time spent in each frequency range in terms of # of generations						
p range	0.001-0.01	0.01-0.1	0.1-0.5	0.5-0.9	0.9-0.99	0.99-0.999
$w_{11}>w_{12}>w_{22}$	462	480	439	439	480	462
dominance, $w_{11}=w_{12}>w_{22}$	232	250	309	1.020	9.240	90.231

55. Another view of the same sort of data. Here we use a conventional notation *s,* for '*selection coefficient'*. The parameter *s* is simply the ratio $1-w_2/w_1$ – e.g., if 100 A's survive compared to 90 a's, then the ratio is 1-0.0 = 0.1. Values of *s* of 0.1 or greater are relatively unsual biologically.

Selection coefficient, *s*	time (generations)
1.0	13.26
0.5	22.67
0.2	50.41
0.1	96.42
0.05	188.36
0.02	464.09
0.01	923.61
0.001	9194.83

56. The times show that while selection ultimately destroys variation, the times are much much longer than those under blending inheritance. We might expect to observe considerable genetic polymorphism in populations even though they are subject to directional selection.

57. Suppose that the fitnesses of *AA, Aa*, and *aa* stand in the ratio $(1+s)^2 : 1+s : 1$. *s* is called a selection coefficient. [NB this is geometric mean – multiplicative fitnesses] Note: 1+*s:* 1 is not the same as 1: 1-*s*. (Close, but no cigar:) The curve of gene frequency change wrt generation time is a logistic curve (see Populus simulation). The time taken to change between any two gene frequencies is (approximately) inversely proportional to *s*.

58. Overdominance. When the fitness of the heterozygote is higher than that of either homozygote, natural selection will bring the gene frequency toward an interior equilibrium, retaining both alleles. This is a polymorphism. The exact equilibrium gene frequency depends on the fitnesses (in fact, if fitnesses are written as $1-s : 1 : 1-t$ the equilibrium frequency of *A* is $t/(s+t)$.[Show this]) If we plot fitness against gene frequency we get a quadratic curve, with a peak precisely at this equilibrium gene frequency.

59. This movement of gene frequencies can be rationalized in terms of the mean fitness of *A* compared to the mean fitness of everybody. In an overdominant case, when *A* is rare it is present mostly in heterozygotes. In that case *A* copies have a higher mean fitness than a (which, being common, are mostly located in aa individuals. So *A* increases when rare. When *A* is common and *a* is rare, the argument is reversed, with *a* being mostly in heterozygotes and having the advantage.

60. Underdominance. When *s* and *t* are both negative, so that the heterozygote is the worst genotype, the gene frequency will move continually away from the interior equilibrium, which is now an unstable equilibrium. The gene frequency tends toward 1 or 0. Which one it goes to depends on which side of the unstable equilibrium it started from. Note that the outcome depends on the exact starting point. The plot of fitness against gene frequency is again a quadratic curve, but now it has a minimum at the unstable equilibrium. The stable equilibria are now 0 and 1.

61. Selection and fitness. In all of these cases the gene frequency changes so that the mean fitness either improves or remains the same, it never declines. In each case the population "climbs" the adaptive surface or fitness surface until it comes to rest at the top. Incidentally, this is true for constant relative fitnesses, and for any number of alleles. It is not perfectly true when fitness is controlled by multiple loci. But in a lot of cases it is true that there is a net gain of mean relative fitness from the beginning of the evolution of the gene frequencies to the end.

62. Is "all for the best in this best of all possible worlds?" (At least in terms of evolution resulting in optimal organisms). The underdominance case shows that while evolution at a single locus (with constant relative fitnesses) results in improvement of the mean fitness, the population can sometimes come to rest on an equilibrium which is not the highest possible one. It depends on the starting point. A gene is evaluated by natural selection against the backgrounds in which it occurs, and that decides whether it will increase. If the fitnesses of *AA, Aa,* and *aa* are 1.2 : 0.7 : 1, then when *A* is rare it is mostly occurring with *a*'s in heterozygotes, which have fitness 0.7. By comparison, the *a*'s are occurring in homozygotes which have fitness 1. So *a* seems to be better and copies of it survive and reproduce better. But in fact, *AA* would be the best genotype. However natural selection is not making a global assessment of the effects of combining alleles, so it misses this and we end up with *aa*. Thus the opportunistic nature of natural selection causes us to climb the nearest peak on the adaptive surface, not the highest one.

63. Thus the opportunistic nature of natural selection causes us to climb the nearest peak on the adaptive surface, not the highest one. If one could always do the latter, would we be able to fly (unaided) at 500 miles per hour, swim to the depths of the ocean, while composing brilliant sonatas all the time? There does not seem to be any way to know, without a comprehensive understanding of organisms.

64. Returning to our 'variance' equation…The basic principle of evolution by natural selection can then be stated as follows: **allele frequencies change in such a way as to maximize the mean fitness $\bar{w}$ of the population**. It is as if ln $\bar{w}$ plays the role of a potential in physics. And it at first looks like we're done – or close to finding a kind of "Hamiltonian" for biology. Or are we??

65. So here comes the big BUT. The equation above turns out to be <u>inadequate</u> as a model of selection in all but the <u>simplest</u> case of independent loci, frequency independent selection, arbitrarily large populations. In reality, neither the maximization of fitness, nor the simple relation between variance (heterozygosity), $p(1-p)$, is realized in fact. Only in the case of a single locus with constant genotypic fitness is this an adequate representation of selection. In all other cases there is neither a necessary maximization of fitness – it may even be minimized! – nor does selection operate most rapidly at intermediate frequencies. So there is no Fermat principle of 'shortest distance' – evolution by natural selection won't look like physics of the sort we're used to. We will use the case of frequency-dependent selection (next lecture) to illustrate; but even the case of 3 alleles for one gene (below, in the sickle cell anemia example), and constant fitnesses, we get the same result.

66. We next immediately generalize to the multi-allele case. This simply involves converting the 2 x 2 matrix of *w*'s we had earlier to general *n* x *n* form ***w.*** We now derive the multidimensional analog of the 'adaptive landscape' equation, as well as the correct model for the sickle-cell anemia case, in which there are 3 common alleles: HbA, HbC, and HbS. *S* homozygotse have sickle-cell anemia, which occurs when the hemoglobin forms long crystals under low oxygen tension. The table below is from Hartl and Clark (1989, p. 171). It gives the observed genotypic counts and Hardy-Weinberg expectations for all six genotypes from a sample of 32,898 individuals from 72 West-African populations:

xxxx	*AA*	*SS*	*CC*	*AS*	*AC*	*SC*
Observed	25374	67	108	5482	1737	130
Expected	25616	307	75	4967	1769	165
Obs/Exp	0.99	0.22	1.45	1.10	0.98	0.79
Relative fitness	0.89	0.20	1.31	1	0.89	0.70

67. Calculating fitness estimate here from H-W (assuming that H-W ratios would be a population with relative fitnesses all equal… is this correct?)

68. Note first that if a population composed entirely of *AA* genotypes, with mean fitness 0.89 was invaded by a single *S* allele (in a heterozygote), *S* would increase in frequency – a single *S* has a fitness of 1.0. With only these two alleles, we would evolve to the (see Populus simulation) equillbrium value for *S* as follows (actual average value over Africa of 0.09)

$$\hat{p}_S = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})} = \frac{0.2 - 1.0}{0.89 - 2.0 + 0.2} = 0.1209$$

69. The mean fitness at equilibrium is 0.9033. [Compute this!]

70. If a second mutation introduces the *C* alleles into a population at equilibrium between *A* and *S,* its spread will be determined by its marginal fitness, which is:

$$w_C = p_A w_{AC} + p_S w_{SC} + p_C w_{CC}, \text{ when } p_c \approx 0,$$
$$w_C = p_A w_{AC} + p_S w_{SC} = 0.8670$$

71. When *C* is rare, p_C is approximately 0 so the third term can be ignored. So, p_A=0.8791, and the marginal fitness of *C* when rare is: 0.8670, less than the mean fitness of 0.90. So, *C* cannot invade when rare, even though a population fixed for *C* would have a global maximum mean fitness. If *C* were introduced in sufficient numbers to include a contribution from the third term, then *C* would fix.

72. It is even more useful to figure out this 'adaptive topology' by looking at each *pair* of alleles at a time: A-S (this we did), C-S, and A-C. This reveals the existence of a 'saddle point' where movement in one direction lowers fitness, and in the other, raises fitness.

73. Does Darwinism mean survival of the fittest?

Biology is not physics. Despite the superficial lure of a Fermat-type maximization/minimization principle, this is not realized.