Contents

1 Evolutionary Thinking

Introduction (Lecture)

- Essential questions of evolutionary biology:
  - Why do organisms look so different?
  - Why develop elaborate sexual traits?
  - Why do organisms senesce?
- Evolution is mainly an historical science and thus must relay on other methods of reconstructing the past or making inferences about evolutionary forces.
- **Proximate**: a question about a mechanistic cause; provides an immediate explanation about how a mechanistic cause functions.
- **Ultimate**: why, or the reason, a trait or organism is the way it is; an evolutionary explanation.
- Example of proximate vs ultimate in Galapagos finches:
  - Proxmate: developmental growth factor is increased/decreased in some birds.
  - Ultimate: different habits are selected on breaks that maxmize food gathering ability.
- Evolutionary biology's approach to answering questions:
  - **Empirical data**: observation studies, experiments; the *comparative method*.
  - **Theory**: predictions that use models and mathematical reasoning which can be be tested with empirical data.
- Overview of the components of evolution by natural selection:
  - Genetic variation exists, via mutations.
  - Mutations are heritable.
  - The is an advantage to survival and/or reproduction from the mutation.
  - Individuals with the advantage in survival/reproduction are selected for.

10 Studying Adaptation

Hypothesis Testing: Oxpeckers Reconsidered

- **Adaption**: a trait, or a suite of traits, that increases the fitness of its possessor.
- No hypothesis for the adaptive value of a trait should be accepted simply because of its plausibility.
- Oxpeckers and impalas traditionally were thought to have a mutually beneficial existence; oxpeckers ate ticks and impalas provided a safe environment.
- Experiments on cattle were done to test whether this observation was true:
  - Results show red-billed oxpeckers have no effect tick loads of cattle.
  - Red-billeld oxpeckers maintained open wounds, even enlarging existing wounds to feed on the cattle's blood.
  - Red-billed oxpeckers removed hosts' earwax; whether this is good of bad is unclear.
  - Even these results must remain in question, as cattle are not the native host for the birds.
- Other important points to remember:
  - Differences among populations or species are not always adaptive.
  - Not every trait is adaptive.
  - Not every adaptation is perfect, often the adaption just happened to work well enough or by chance better than other adaptations.

Experimental Design

- Defining and testing effective control groups is critical.
- Treatments of controls and experimental measures must be handled as close to exactly alike as possible.
- Randomization is a key technique for equalizing miscellaneous effects and a tool to avoid bias.
- Reproduction is essential in order to help remove potential outlier effects.
  - Allows for greater understanding of precision, accuracy, and variation by providing more data for statistical tests.

4 Evolutionary Trees

How to Read an Evolutionary Tree

- **Phylogeny**: aka evolutionary tree or phylogenetic tree, is a diaggram showing the history of divergence and evolutionary change. Essentially, it's the genealogical relationships of organisms based on descent with modification.
  - **Taxa**: the units you are analyzing, e.g. certain species or DNA sequences.
    - **Character**: a feature or trait present among the taxa of interest, e.g. teeth of mammals or nucleotides of DNA sequences.
      - *Character state(s)*: an alternative condition of a character, which are able to evolve one to another, e.g. pointed/flat teeth of mammals.
    - *Ancestral character*: a trait that was possessed by the common ancestor.
    - *Derived character*: a trait the was not possessed by the common ancestor and instead evolved in at least one of the descendants.
      - **Synapomorphy**: derived character state shared by two or more taxa and used to define a clade of taxa.
      - **Autapomorphy**: derived character state in only one taxon.
    - **Outgroup**: a taxon or taxa that are used to root the phylogeny or determine ancestral character states.
    - **Ingroup**: the set of taxa that are the focus of the phylogeny.
  - **Nodes**: points at which the tree splits; represents mutations, speciation events, or character changes.
  - **Anagensis**: descent with modification, but no speciation.
  - **Cladogenesis**: speciation, origin of clades.
    - **Clade**: also known as a monophyletic group, an ancestor and all of its descendants.
    - **Paraphyletic group**: a group of organisms consisting of an ancestor and some of its descendants.
  - **Sister**: a taxa or clade that are most closely related to each other; they share the most recent common ancestor.
- **Homology**: similarity due to common descent; continuity of a trait, character, or character state through time.
  - *Homologous trait*: found in a taxa that inherited the trait from a common ancestor.

- **Homoplasy**: or analogous, similarity in the characters or traits in different taxa due to convergent evolution, parallelism, or reversal, but not due to common descent.
  - **Convergent evolution**: similar traits due to selective forces and not shared ancestry.
    - **Parallelism**: convergent evolution in recently diverged taxa.
  - **Reversal**: derived traits or character states that revert to the ancestral form.

Inferring Phylogenetic Trees

- **Parsimony**: relationships that require the smallest number of character changes are most likely to be correct.
  - Based on shared and derived traits(synapomorphies).
  - Reconstruction using parsimony:
    1. Code characters.
    2. Make up a taxon×character matrix.
    3. Search for synapomorphies, and the shortest tree.
    - Outgroups can help polarize (ancestral vs derived) the characters.
  - **Treelength**: a measure of evolutionary change using parsimony.
    - Shortest tree length produces most parsimonious tree.
    - Length determined by number of synapomorphies.
    - Homoplasious characters increase tree length.
- **Distance Methods**: converts a sequence alignment to genetic distances between pairs of sequences.
  - Branch length is proportional to genetic differences.
- **Maximum likelihood**: a method of estimating the parameters of a probability distribution by maximizing a *likelihood function*.
  - One of the more dominant means of statistical inference.
  - **Likelihood**: measure of goodness of fit of a statistical model to a sample of data for given values of the unknown parameters.
  - $P(D|H)$; probability(P), Data(D), Hypothesis(H)
  - **Bayesian**: uses the likelihood function to create a quantity called the *posterior probability* of trees using a model of evolution based on prior probabilities in order to produce the most likely tree.

- **Bootstraping**: creating a value that indicates how many times out of 100 (normally) that the same branch was observed when repeating the phylogenetic reconstruction on re-sampled (pseudoreplicated) set of dat.

6 Mechanisms of Evolutionary Change

Hardy-Weinberg Equilibrium

- **Population**: a group of interbreeding individuals and their offsring.
- **Gene pool**: the set of all genes, or genetic information, in any population.
- **Genotypic frequency**: number of individuals with a given genotype divided by the total number of individuals in the population.
  - The proportion (i.e., $0 < f < 1$) of genotypes in a population.
- **Allele frequencies**: relative frequency of an allele at a particular locus in a population.
  - **Locus**: a fixed position on a chromosome where a particular gene of genetic marker is.
  - Monoploids: frequency of an allele is the result of the number of copies of the allele divided by sample size.
    - $p = i/N$
    - p: frequency | i: copies of alleles | N: sample size
  - Diploids: frequency of alleles within three possbile genotypes at a locus with two alleles.
    - $p = f(AA) + \frac{1}{2}f(AB)$ frequency of A-allele
    - $q = f(BB) + \frac{1}{2}f(AB)$ frequency of B-allele
  - Allele frequency can always be calculated from genotype frequency, whereas the reverse requires the *Hardy-Weinberg principle* of random mating apply.
- **Hardy-Weinberg principle**: allele and genotype frequencies in a population will remain constant in the absence of evolutionary influences.
  - Allele frequencies do not change from one generation to the next.
  - Genotypic frequencies after one generation of random mating: $p^2 + 2pq + q^2$
  - Evolutionary influences: genetic drift, mate choice, assortative mating, natural selection, sexual selection, mutation, gene flow, meiotic drive, genetic hitchhicking, population bottleneck, founder effect, and inbreeding.
    - *Most of these influences will be discussed later.*

Selection

- **Fitness**: success at which a organism produces fertile offspring.
- **Competition**: an interaction between organism in which the fitness of one is lowered by the presence of another.
- **Selection**: the act on a heritable phenotypic trait due to competition.
  - Can be members of the same of different species.
  - Not always directional and adaptive, instead selection pressure is applied and removes the less fit variants.
  - Can be classified in different ways, such as effect on a trait, on genetic diversity, by life cycle, by unit of selection, or by the resource in competition.
  - Most effective on large populations.

By Effect on a Trait

- **Stablizing selection**: the simplies case in which selection acts to hold a trait at a stable optimum.
- **Directional selection**: favours extreme values of a trait.
- **Disruptive selection**: acts during transition periods when current mode is sub-optimal, but alters trait in more than one direction.
  - **Univariate**: when the trait is both quantitatively favoured in either direction and can lead to speciation.

By Effect on Genetic Diversity

- **Purifying selection**: aka negative selection; acts to remove genetic variation from the population.
- ***de novo* mutation**: introduces new variation and opposes negative selection.
- **Balancing selection**: acts to maintain genetic variation, even in absence of *de novo* mutation by frequency-dependent selection.
  - **Frequency-dependent selection**: fitness that depends of the phenotypic or genotypic composition of a population.
    - Positive: fitness increases as frequency of the trait increases.
    - Negative: fitness decreases as the frequency of the trait increases.

- **Overdominance**, *aka heterozygote advantage*: when a combination of alleles confers a selective advantage over individuals with one allele.
- **Underdominance**, *aka heterozygote disadvantage*: when the heterozygote has lower fitness than either homozygote.

By Life Cycle Stage

- **Viability selection**: *aka survival selection*: increases probability of survival.
  - Can act to improve probability of survival before and after reproduction.
- **Fecundity selection**: increases the rate of reproduction given survival.
  - May be split into sub-components including sexual selection, gametic selection, gamete viability, compatability selection, and zygote formation.

Mutation

▷ **Mutation**: alteration in the nucleotide sequence of the genome of an organism.

- May not produce discernible phenotypic changes.
- The ultimate source of genetic variation.
- Have several types of changes, from no effect, to small changes, or complete loss of function.

Large-Scale Structural Mutations

- **Gene duplications**, *aka amplifications*: repetition of a chromosomal segment or attachment of extra piece of chromosome to another, leading to multiple copies of chromosomal regions.
- Deletions of large chromosmal regions.
- **Fusion genes**: mutations that join previously separated genes into one new distinct gene.
- **Chromosmal rearrangement**: large scale changes in structure of chromosomes, leading to speciation in isolated, inbred populations. Includes:
  - **Chromosomal translocations**: interchange of genetic parts from nonhomologous chromosomes.

    - **Chromosomal inversions**: reversing the orientation of a chromosomal segment.
    - Non-homologous chromosomal crossover.
    - **Interstitial deletions**: inverse of fusion genes; removes a segment of DNA joining distant genes.
  - **Loss of heterozygosity**: loss of one allele, by deletion or genetic recombination, in a organism that previously had two different alleles.

Small-Scale Mutations

  - **Point mutation**: a single nucleotide base change, that can result in a variety of effects.
  - **Insertions**: add one or more extra nucleotides into the DNA.
    - Usually caused by transposable elements, or errors during replication or repeating elements.
    - Can cause*reading frame shift*, possibly effecting how many codons are read, and thus altering the gene product.
  - **Deletions**: remove one or more nucleotides from the DNA.
    - Also can cause a reading frame shift like insertions.
    - Generally irreversible.
  - **Substitutions**: exchange of a single nucleotide for another.
    - Often classified as transitions or transversions.
    - Generally a purine (A-G) for a purine, or a pyrimidine (C-T) for a pyrimidine.
    - Can be reversed by another point mutation.

Impact on Protein Sequence

- Effect of mutation depends heavily on where it occurs, particularly in a coding or non-coding region.
- Regulator sequences, e.g. promoters, enhancers, silencers, can alter gene expression but are less likely to alter protein sequence.
- **Frameshift mutation**: caused by insertion or deletion of nucleotides that is not divisible by three, resulting in a different translation from the original.
- **Synonymous substitution**: a condon replacement with another that codes for same amino acid.
  - **Silent substitution**: no phenotypic difference after a synonymous substitution.
- **Nonsynonymous substitution**: a codon replacement that codes for a different amino acid.
  - **Missense mutation**: codon replacement that renders the resulting protein nonfunctional.
  - **Nonsense mutation**: codon replacement that results in a premature stop codon that produces a truncated and often nunfunctional protein.

Migration

- **Gene flow**: movement of alleles, or genetic variation, between populations.
  - If the rate of gene flow is high enough, then two populations are considered to have equivalent allele frequencies and thus a single population.
  - Constrains speciation by combining gene pools of the groups.
  - May result in the addition of novel genetic variants in the gene pool.

Factors of Gene Flow

- Gene flow is expected to be lower in species that:
  - have low mobility or dispersal.
  - occur in fragmanted habits.
  - have long distances between populations.
  - have small population sizes.
- **Allopatric speciation**: when gene flow is blocked by physical barriers that inhibit gene flow.
- **Sympatric speciation**: result of gene flow that is blocked due to non-physical barriers that inhibit gene flow.
  - Often reproduction barriers are the main factors.

Genetic Drift

- **Genetic drift**: the change in the allele frequencies in a population due to random sampling.
  - Not influenced by environmental factors.
- May cause certain gene variants to become fixed or lost by chance.
- Generally drives populations towards genetic uniformity over time, decreasing heterozygosity.
- Only mutation or gene flow can introduce new alleles, which acts against genetic drift.
- **Founder effect**: result of sampling error which has an increased likelyhood on populations with low numbers.
  - By chance certain alleles can be dominant when they otherwise wouldn't be in a new founding population.
  - Often acts to drastically increase rate of genetic drift.
- **Genetic bottleneck**: a sharp reduction in the size of population due to environmental events.
  - Can essentially cause a founder effect, though it's not a new population.

Coalescent Theory

- **Coalescent theory**: how gene variants sampled from a population may have originated from a common ancestor.
  - Assumes no recombination, no natural selection, no gene flow in the simpiliest case.
- Aims to look backward in time by merging allels into a single ancestral copy according to a random process in coalescence events.
- Many theoretical genealogies are made in order to compare to observed data in order to test assumptions about demographic history of a population.
  - Used to make inference about population genetic parameters, such as migration, population size, and recombination.
- **Coalescent time**: number of preceding generations where the coalescence took place, not calender time.
  - Estimation of the time can be made multiplied by $2N_e$ with the average time between generations.
  - Time to coalescence for a pair of allels at a locus is dependent on population size.
  - Formula: $P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1}\left(\frac{1}{2N_e}\right)$
- Can also be used to model the amount of variation in DNA sequences expected from genetic drift and mutation.

Molecular Evolution

- **Molecular evolution**: the process of change in the sequence composision of cellular molecules across generations.
- **Polymorphism**: occurrence of two of more clearly different morphs, or alternative phenotypes, in the population of a species.
  - **Substitution**: when allels become fixed or lost in a population and polymorphism is ended.
    - Substitution rate (k): $k = sN\mu$
    - s = probability of fixation.
    - $N\mu$= mutation rate of population.

Recombination

- **Recombination**: the process that results in genetic exchange between chromosomes or chromosomal regions.
  - Can also cause mutations due to misalignment after recombination.
  - **Gene repair**: a type of recombination that is the product of DNA repair that corrects damage using a homologous template.
    - Often responsbile for homogenizing sequences of duplicate genes over long periods of time, which reduces nucleotide divergence.
- **Genetic hithicking**: change in allele frequency not because of natural selection, but due to proximity to a gene undering selective sweep.
  - **Selective sweep**: a beneficial mutation that increases frequency and generally becomes fixed.

Neutral Theory

- **Neutral theory of molecular evolution**: most evolutionary changes occur at the molecular level.
- Most variation is due to random genetic drift of mutant alleles that are selectively neutral.
- Compatable with phenotypic evolution, as phenotypes are driven by molecular changes.
- Most mutations are neutral with respect to fitness.
- A minority of mutation are advantageous.
- Substitution rate predicted to be neutral, equal to per-individual mutation rate, independent of population size.
- K_A/K_s test used to determin direction selection based on evolutionary history.
  - K_A: number of nonsynonymous substitutions (replacement).
  - K_S: number of synonymous substitutions (silent).
  - $K_A > K_s$ signals for positive selection.
  - $K_A < K_s$ signals for purifying selection.

Molecular Clocks

- **Molecular clocks**: the average rate at which species' genomes accumulates neutral mutations over time.
  - A linear rate is often easy to establish.
  - Used to measure evolutionary divergence.

Nonrandom Mating

- **Inbreeding**: production of offspring from closely genetically related individuals.
  - Results in homozygosity, which can increase chances chances of offspring being affected by deleterious or recessive traits.
  - **Inbreeding depression**: the reduced fitness in a given population due to inbreeding.
    - Usually caused by population bottlenecks or the founder effect.
  - Can also result in purging of deleterious allels through purifying selection.
  - Can allow for the expression of advantageous phenotypes, which if outweighs the disadvantages, then could potentially lead to speciation.
  - **Coefficient of inbreeding**: the probability that two alleles at any locus in an individual are identical by descent.
  - Nonrandom mating does not alter allele frequencies and not a mechanism of evolution.
    - Can alter the frequencies of genotypes, changing the distribution of phenotypes in a population, which can alter patters of natural selection.
- **Assortative mating**: mating based on phenotypic factors.
  - Can play a role in sympatric speciation.
  - A form of sexual selection.
  - Can be either positive of negative, selecting for similar or different phenotypes respectively.

8 Evolution at Multiple Loci

Linkage Equilibrium and Diesquilibrium

- **Linkage equilibrium**: when the genotype of a chromosome at one locus is independent of its genotype at another locus.
- **Linkage disequilibrium**: the non-random association of alleles at different loci in a given population.
  - Occurs when frequency of the association between loci's different alleles is higher or lower than expected.
- **Haplotype**: a group of alleles in an organism that are inherited together from a single parent.
  - Used to mean the collection of specific alleles that represent a phenotype and likely to be conserved.
  - Also can be used to mean a set of linked single-nucleotide polymorphism alleles that are associated statistically.
- Factors that influences disequilibrium: selection, rate or genetic recombination, mutation rate, genetic drift, system of mating, population structure.
  - Undertanding linkage desequilibrium in a genome can be a powerful signal of the population genetic processes that structure it.
  - Selection, genetic drift, assortative mating, and population admixture act to create disequilibrium.
  - Recombination and outbreeding act to reduce disequilibrium.
- Level of linkage disequilibrium can between A and B can be quantified by the coefficient of linkage disequilibrium, D_{AB}.
  - Formula: $D_{AB} = P_{AB} - P_A P_B$
  - P_{AB}: the frequency with which both occur together on same gamete, or the frequency of the AB haplotype.
  - $P_A P_B$: product of the probabilities give the probability they occur together.
  - When there is a difference, the magnitude of the coefficient rises, indicating linkage disequilibrium.
  - Strong recent selection can be indicated by linkage disequilibrium of allels located next to neutral allels.

Recombination's Effect on Linkage Diesquilibrium

- Linkage diesquilibrium (D) will converge to zero depending of the magnitude of the recombination rate (c) between two loci the absence of natural selection, inbreeding, and genetic drift.
- The smaller the distance between the two loci, the smaller the rate of convergence of D to zero.
- Genetic recombination tends to randomize genotypes, thus it tends to reduce frequency of overrespresented chromosome haplotypes and increases underrepresented haplotypes.
- In short, recombination reduces linkage disequilibrium.

Adaptive Significance of Sex

▷

9 Quantitative Genetics

The Nature of Quantitative Traits

- **Quantitative genetics**: deals with phenotypes that vary continuosly due to multilocus traits and environmental factors.
  - Allows for the prediction of how a population will respond to selection, even when we do not know the genetic basis of the trait.
- **Qualitative traits**: descrete traits that can be assigned individual categories by observation of simple genetic test.
- **Quantitative traits**: continuous traits determined by the combined influence of the genotype at multiple loci and the environment; the focus of quantitative genetics.
  - Study of continuous distribution requires many other statistical methods such as the effect size, mean, and variance, to link phenotypes to genotypes.

Heritable Variation

- **Heritability**: the degree of variation in a phenotypic trait in a population that is due to genetic variation betwen individuals in that population.
- **Phenotypic variance (V_P)**: the genetic variance (V_G) combined with the environmental variance (V_E).
  - $V_P = V_G + V_E + V_{GE}$
  - V_{GE} represents variance associated with intereaction of genetic and environmental factors.

Genetic Variance

- **Additive genetic variance (V_A)**: how much the phenotypic trait is influenced by traits that show an additive effect on the quantitative traits.
  - Measures the magnitude to which individual phenotypic differences can be predicted due to additive effects of allelic substitutions.
- **Dominance genetic variance (V_D)**: associated with the dominant gene actions which cover the influence of the recessive alleles at the particular locus.
- **Epistasis (V_I)**: occurs due to statistical interaaction among loci, i.e., gene-by-gene modification.
- $V_G = V_A + V_D + V_I$

- **Broad-sense heritability (H^2)**: all genetic contributions to a populations phenotypic variance, including additive, dominant, and epistatic, and maternal/paternal effects.
  - $H^2 = \frac{V_G}{V_P}$
- **Narrow-sense heritability (h^2)**: proportion of the total phenotypic varian that is due to the additive effects of genes.
  - $h^2 = \frac{V_A}{V_P}$
  - Allows prediction of how a population will respond to selection.
  - The greater the additive genetic variation for the trait, the greater its response to selection can be.

Measuring Differences in Survival and Reproductive Success

- Heritability is often estimated by measuring correlations between parents and offspring.

Predicting Evolutionary Responses

-

11

▷

12

▷

13

▷

14

▷

16

▷

17

▷

18

▷

20

▷

15

▷