Evolutionary Systems Biology

Instructor: Bradly Alicea
bradly.alicea@outlook.com
Join us on Slack and Discord for office hours and discussion!

Three modules (Genes and Systems, Cells, Organs, and Organismal Self-organization I, and Cells, Organs, and Organismal Self-organization II).

Lecture #1: Genes and Systems.

- what is evolution? Examples of what evolution is and is not, experimental and theoretical approaches.
- interactions between genes (epistasis), synthetic/suppressive effects, and gene action (example of processes that set up evolution).
- graphs and evolutionary representations (example of theory).

Lecture #2: Cells, Organs, and Organismal Self-organization I.

- myth of "junk" DNA? Yes and no (genome content and structure).
- gene networks and biocomplexity (genome function).
- synthetic life and bioengineering (relevance of evolutionary systems biology to regenerative medicine and systems engineering).

Lecture #3: Cells, Organs, and Organismal Self-organization II.

- physiomics and scales of life (combining methods vertically- genes to organism)
- modularity and evolution (functional subdivisions of life)
- specific mechanisms and predictions (epigenetics, phenotypic capacitance, and facilitated variation).

Book and article sources: for general reference and further reading.
*** Bold** = required reading.

General Sources:

Alon, U. (2007). An Introduction to Systems Biology: design principles of biological circuits. CRC Press, Boca Raton, FL.

Bromham, L. (2008). Reading the Story in DNA: a beginner's guide to molecular evolution. Oxford University Press, Oxford, UK.

Carroll, S., Grenier, J., Weatherbee, S. (2004). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Wiley-Blackwell, New York.

Davidson, E.H. (2001). Genomic Regulatory Systems: Development and Evolution. AcademicPress, New York.

Hein, J., Schierup, M.H., and Wiuf, C. (2005). Gene Genealogies, Variation, and Evolution: a primer in coalescent theory. Oxford Press, Oxford, UK.

*** Medina, M. (2005). Genomes, phylogeny, and evolutionary systems biology. PNAS USA, 102(1), 6630–6635.**

Nowak, M. (2006). Evolutionary Dynamics. Harvard, Cambridge, MA.

Rice, S.A. (2007). Encyclopedia of Evolution. Checkmark Books, New York.

*** Nature's 15 Evolutionary Gems. http:///www.nature.com/evolutiongems/**

Specific Topics:

Emergence, Evolvability and recent theoretical developments:
Crombach, A. and Hogeweg, P. (2008). Evolution of Evolvability in Gene Regulatory Networks. PLoS Computational Biology, 4(7), e1000112. Apply GRNs to understanding evolvability.

Flintoft, L. (2007). Developmental networks in time and space. Nature Reviews Genetics, 8, 88-89.

Hoekstra, H.E. and Coyne, J.A. (2007). The Locus of Evolution: evo-devo and the genetics of adaptation. Evolution, 61(5), 995–1016. Basic predictions of evo-devo.

Kirschner, M. (2005). Plausibility of Life: resolving Darwin's dilemma. Yale University Press, New Haven. Primer for facilitated variation - FV.

Muller, G. (2007). Evo-Devo: extending the evolutionary synthesis. Nature Reviews Genetics, 8, 943-949. Review of the evo-devo field, current state-of-the-art.

Stewart, A.D. and Hunt, D.M. (1982). The Genetic Basis of Development. Halsted Press, London, UK.

Wagner, A. (2005). Robustness and Evolvability in Living Systems. Princeton University Press, Princeton, NJ.

Wilkins, A.S. (2002). The Evolution of Developmental Pathways. Sinauer, Sunderland, MA.

Functional Genomics, Phenomics, and Biocomplexity:

*** Adami, C. (2006). Reducible Complexity. Science, 312(7), 61-63.**

Bergman, A. and Siegal, M.L. (2003). Evolutionary capacitance as a general feature of complex gene networks. Nature, 434, 549-552. (phenotypic capacitor review – capacitor selectively releases phenotypic variation via gene expression under certain conditions).

Greally, J.M. (2007). Encyclopedia of humble DNA. Nature, 447, 782-783. (review of ENCODE project; identification of functional elements in human genome).

LeMeur, M. and Gentleman, R. (2008). Modeling Synthetic Lethality. Genome Biology, 9, R135. 1-2 page review of epistasis (interactions in polygenic systems) and synthetic lethality (experiments to uncover gene action when expressed in combination).

Tucker, C.L. and Fields, S. (2003). Lethal Combinations. Nature Genetics, 35(3), 204-205. 1-2 page review of epistasis (interactions in polygenic systems) and synthetic lethality (experiments to uncover gene action when expressed in combination).

Weibel, E.R., Taylor, C.R., and Hoppeler, H. (1991). The concept of symmorphosis: a testable hypothesis of structure-function relationship. PNAS USA, 88(22), 10357-10361.

Experimental Evolution:

Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. PNAS USA, 105, 7899-7906.

Garland, T. (2003). Selection experiments: an under-utilized tool in biomechanics and organismal biology. In "Vertebrate biomechanics and evolution", V.L. Bels, J-P. Gasc, and A. Casinos eds., Pages 23-56. BIOS Scientific Publishers, Oxford, U.K.

Kolss, M. and Kawecki, T.J. (2008). Reduced learning ability as a consequence of evolutionary adaptation to nutritional stress in *Drosophila melanogaster*. Ecological Entomology, 33, 583-588.

Leu, J-Y. and Murray, A.W. 2006. Experimental Evolution of Mating Discrimination in Budding Yeast. Current Biology, 16, 280-286.

Synthetic Biology:

Benner, S.A. and Sismour, A.M. (2005). Synthetic Biology. Nature Reviews Genetics, 6, 533-544. Review of the synthetic biology field.

Endy, D. (2008). Reconstruction of the genomes. Science, 319(5867), 1196-1197. http://www.jcvi.org/cms/research/projects/synthetic-bacterial-genome/press-release/ Reviews of the first customized microbe using a synthetic genome approach developed at JCVI (J. Craig Venter Institute).

*** Endy, D. (2005). Foundations for engineering biology. Nature, 438, 449-453.**

Kauffman, S.A. (1993). The Origins of Order: Self-Organization and Selection in Evolution. Oxford, University Press, Oxford, UK.

Taylor, C. and Jefferson, D. (1995). Artificial Life as a tool for biological inquiry. In Artificial Life: an overview. C.G. Langton, ed. MIT Press, Cambridge, MA. Available via Google Books.

Physiomics, transcriptional/translational regulation:
Barabasi, A-L. and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5, 101-113.

Biemont, C. and Vieira, C. (2006). Genetics: Junk DNA as an evolutionary force. Nature, 443, 521-524.

Clarke, J.D.W. and Tickle, C. (1999). Fate maps old and new. Nature Cell Biology, 1, E103-E109.

Espinosa-Soto, C., Padilla-Longoria, P., and Alvarez-Buylla, E.R. (2004). A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. The Plant Cell, 16, 2923–2939. Features a genetic regulatory network (GRN) model example.

Hunter, P.J. and Borg, T.K. (2003). Integration from proteins to organs: the Physiome project. Nature Reviews Molecular Cell Biology, 4, 237-243.

Kitano, H. (2006). Computational cellular dynamics: a network–physics integral. Nature Reviews Molecular Cell Biology, 7, 163.

Markowetz, F. and Spang, R. (2007). Inferring cellular networks: a review. BMC Bioinformatics 8(6), S5.

Matzke M.A. and Matzke, A.J.M. (2004). Planting the Seeds of a New Paradigm. PLoS Biology, 2(5), e133.

Matzke, M. and Matzke, A.J.M. (2003). RNAi Extends Its Reach. Science, 301(5636), 1060-1061.

Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics, 23(3), 314-318.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature, 428, 717-723.

Waddington, C.H. (1953). Genetic Assimilation of an Acquired Character. Evolution, 7(2), 118-126.

Homework:
Two components: 1) phylogenetics and 2) models of heredity and phenotype.

Objectives for phylogenetics: understand the context in which a tree is an appropriate way to think about evolution, build a phylogenetic tree, become familiar with some of their features, and become acquainted with several famous applications of phylogeny.

Objectives for heredity and phenotype: understand the basic logic and linear algebra of a single locus, two-allele model of Hardy-Weinberg equilibrium, understand the effects of gene action on a phenotype, extrapolate this model to a two locus, two-allele situation (e.g. the Agouti mouse phenotype).

Phylogenetics:

Introduction. The objective of this assignment is to understand the basics of phylogenetic relationships and why they are important in understanding the patterns of biodiversity observed in nature.

Scientists as far back as Darwin and Linnaeus conceived of natural variation as being "tree-like". This implies that all existing diversity originated at a single point in the past and eventually split off into a series of related but non-recombining categories. Up until the early 20th century, the tree of life was thought to be progressive, with humans and other so-called "higher" animals being the ultimate outcome of evolution.

Q1: what is incorrect about a progressive view of evolution? Can there be exceptions to the non-reticulating (e.g. tree-like) model of evolution? HINT: read through assignment or do additional research via Google before answering. 3 points.

Cladistics and Phylogenetics. In the 1960s Willi Hennig introduced a method called cladistics which does not assume that different biological categories are higher or lower, just more basal or more derived (in fact, the terms basal and derived are the correct way to refer to a group of related species: species that presumably diverged early in the history of a taxonomic group are called basal, while those that diverged later in time are derived). This approach is known by the more well-known term of phylogeny, and organizes biodiversity in a number of unique ways:

1) traits are identified as being either ancestral or derived. Given a common ancestor, various states of a trait can be traced to a common origin point.

2) each nested group is called a clade, which defines a number of biological units as being closely related. Clades are generally defined by synapomorphies, or traits (characters) that are both shared and derived.

3) derived traits shared by more than one species are either homologous or homoplastic:

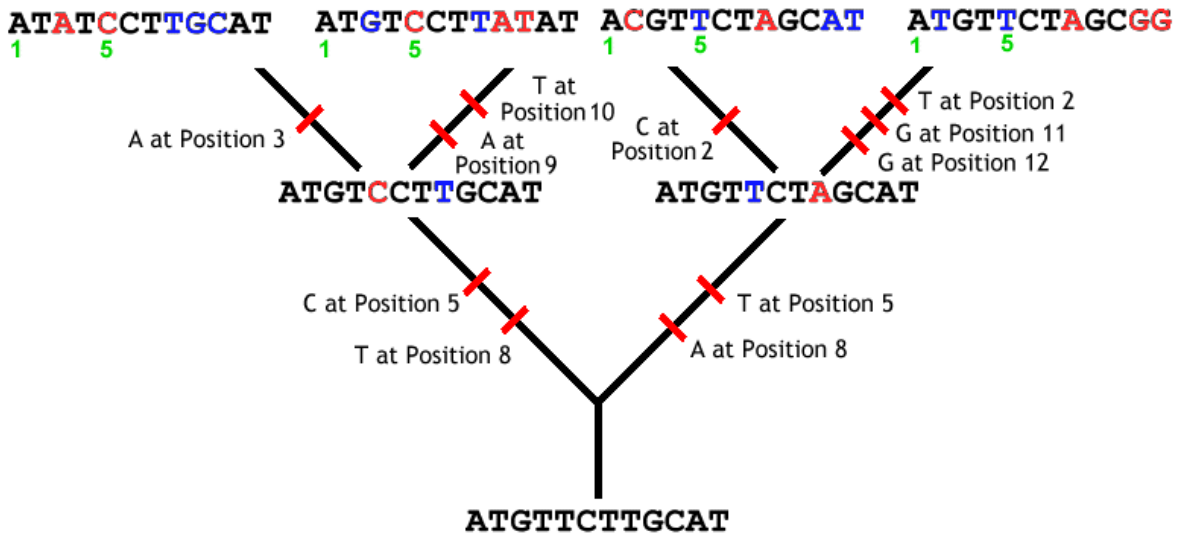
Examples of Homology	Examples of Homoplasy
All bird wings	Bird wings vs. Bat wings
Linkage within a specific gene family	Gene function in different gene families
Gene sequence in related species	Gene sequence similarity in genome-wide context

4) the shape of the phylogeny is defined by a series of evolutionary steps, each of which represents a state change for a single trait. These steps are discretized characterizations of a trait (e.g. different beak shapes are identified as 0, 1, and 2, or different nucleotides at the same base position are referred to as A, T, C, and G).

5) tree structures are rooted using a distantly-related species in which shared characteristics are more likely to be ancestral than derived. In this case, evolutionary conservation plays a role in helping us identify evolutionary relationships. Specifically, this allows us to polarize various traits.

6) different trees are built using different sets of assumptions. One method, called parsimony, assumes that the simplest explanation is the best. Consequently, a tree built using parsimony does not include many losses of derived characters or convergent evolution (homoplasy). Other schools of thought, such as Maximum Likelihood, assume that the evolution of characters involves a likelihood function, as in which character states are most likely to appear in a given taxon.

Hennigian argumentation: how to trace characters through a tree. One approach to building a phylogeny from a limited number of characters (there exists software for building larger trees) is by defining evolutionary relationships using single character state changes. The following gene tree gives a graphical example of this.



As we can see, a "C" in position 5 and a "T" is position 8 defines the clade on the left-hand side of the graphic. The tips of the tree can be defined by additional changes in the sequence. In cases where we have the information (e.g. mutation rate), branch lengths are of variable lengths. Each state change is summarized using hash marks along the branch they are hypothesized to occur. This can also be done with binary character states (e.g. zeros and ones).

Q2: build a phylogeny according to the following matrix:

Species	Trait1	Trait2	Trait3	Trait4	Trait5	Trait6
A	0	0	0	0	1	1
B	0	1	0	1	0	0
C	0	1	1	0	0	0
D	0	0	0	0	1	0
E	1	0	0	0	0	0

Use the Hennigian argumentation approach (link). Each branch should contain at least one state change. Assume branch lengths are equal. HINT: species in the same clade might share all but one trait. 5 points.

List all synapomorphic traits. 3 points.

How many derived states exist in taxon A? 3 points.

How many ancestral traits are **shared** by taxa C and D? 3 points.

A few problems with this approach:

A) trait recombination: especially in the case of neural characters, evolution can use a number of solutions to arrive at the same trait shape, state, or function. In these cases, the traditional "tree of life" model is violated, and one cannot tell (especially in light of other evidence) whether a character state is homologous or not.

B) long-branch attraction: when two species share a large number of derived traits, it is hard to tell how long ago they diverged.

C) unresolved groups: there is not enough unique information in each species to tell them apart using a phylogenetic approach. These relationships are represented in the phylogenetic topology by a series of parallel lines.

D) homoplastic traits: derived traits that are shared by species in different clades. Often times this is called convergent or parallel evolution.

Q3: you are asked to build phylogeny out of the following matrix:

Species	Trait1	Trait2	Trait3	Trait4	Trait5	Trait6
A	0	1	1	1	1	1
B	0	1	1	1	1	1
C	0	1	1	0	0	1
D	0	0	1	1	1	0
E	1	1	1	0	0	0

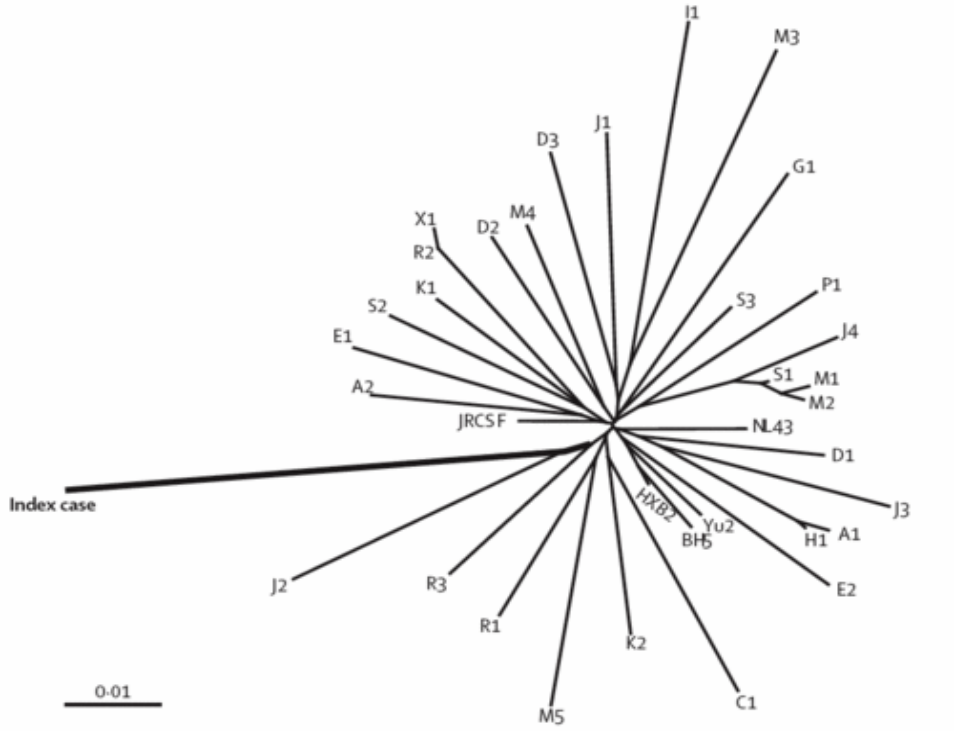
Build a phylogeny as you did in #2. Identify instances of A through D in the resulting topology. 5 points.

Instances of A through D. 3 points.

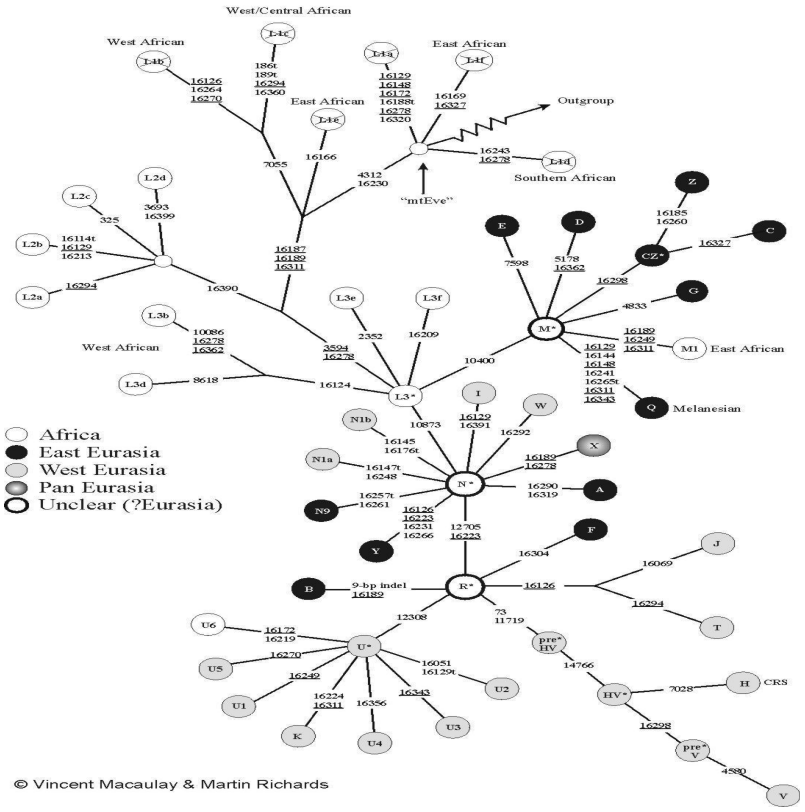
Extra credit: Identify famous phylogenies.

Look at the following phylogenies and identify what systems they are representative of and/or what's going on in each example (pay attention to clues in the figure and search Google).

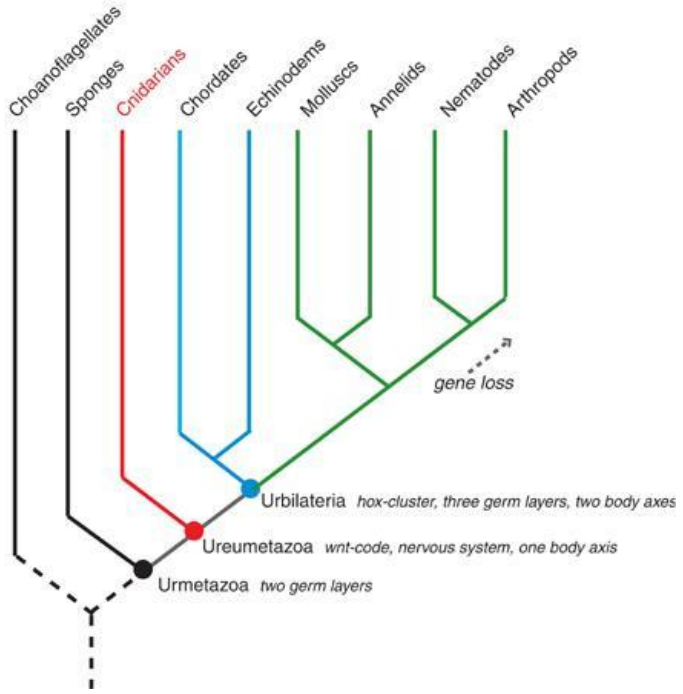
1. Hint: evolution of a modern scourge (viral evolution). 3 points.



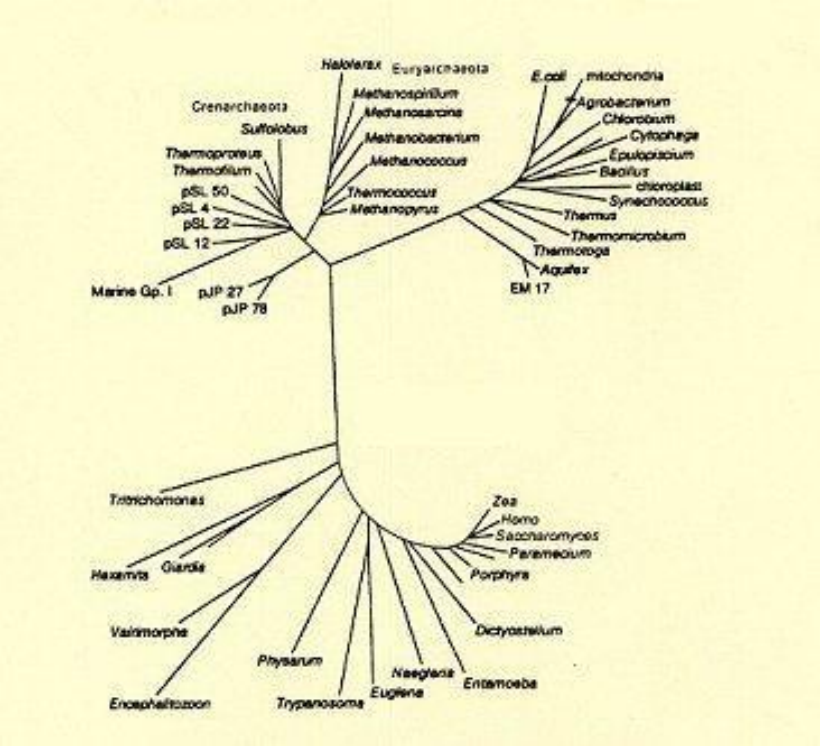
2. Hint: regional variation in a non-Primate group (within a single species). 3 points.



3. Hint: all animals are as one (least recent common ancestor). 3 points.



4. Hint: life's diversity is larger than previously thought (three major groups). 3 points.



Models of Heredity and Genotype

* you may show your work in either equation form or Matlab/Python code

The Hardy-Weinberg model is a way to model a genotype from observations of a phenotype, or the probability of the offspring of two parents as either carriers or expressers of a particular genotype. This is useful in the medical domain, where genetic counselors can predict how likely a child is to develop a certain disease.

Deriving a phenotype from a genotype is generally a complex affair. For example, most traits are polygenic, epistatic (genes that interact during expression), and are influenced by environmental factors. However, as a conceptual and predictive tool, we can use relatively simple models of "hard" inheritance to understand this relationship.

Law of Independent Assortment

Gregor Mendel, in his experiments with peas, discovered a number of principles regarding genotypic distributions based on observations of phenotype across generations. Keep in mind that this was done before the concepts of the gene and DNA were made concrete. Nevertheless, it is one of the cornerstones of the modern evolutionary synthesis. One of these principles is the **Law of Independent Assortment.** This law predicts that each parent (**F_1 generation**) carries two copies of a certain locus, and that during **meiosis** (sexual reproduction) the four copies of this locus have an equal chance of being transmitted to the offspring. The offspring (**F_2**

generation) will inherit only two of these four copies, so that multiple offspring from the same set of parents will have different genotypes. Keep in mind that this ignores cases in which certain loci are passed down either maternally or paternally (sex-linked traits).

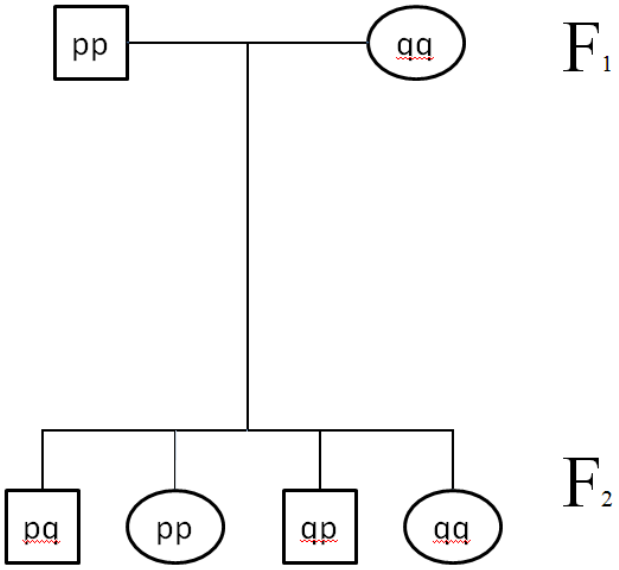
Terminology

A **locus** or **site** is analogous to a gene, but can also represent any DNA sequence in the genome.

An **allele** is a variant locus or site. Generally, an allele represents both a structural (sequence) and a functional (contribution to phenotype) difference from other alleles. Alleles can be **dominant** or **recessive** in their effects.

A **homozygote** is a genotype that inherits that same allele from each parent. A **heterozygote** is a genotype that inherits one allele from one parent and another allele from the other parent. This has consequences for gene action: a **recessive** phenotype requires two copies of the recessive allele for a particular trait, while a dominant phenotype can be derived from either a single or both copies of the dominant allele.

In population genetics, models exist which consider both alleles and sites (loci) to be infinite. For many complex traits, there can be many alleles and many sites all with unique effects on the system.



Standard geneaology showing relationship between F1 (parental generation) and F2 (offspring). Squares are males, circles are females. Note that both parents are homozygous, but the offspring exhibit all possible combinations (homozygous and heterozygous).

Hardy-Weinberg Mechanics (9 points)

Basic Equation

The Hardy-Weinberg technique is at its heart a probabilistic tool. In general, the field of population genetics treats the problem of single-locus heredity as a **binomial expansion**. The expression

$$1 = (p + q)^2$$

tells us that two alleles (*p* and *q*) can be observed as proportions of a phenotype in a population. Let us assume that in a population of birds that *p* is a red beak and *q* is a blue beak. As the equation suggests, these frequencies must equal 1.0. For an F1 population of birds, $p = .30$ and $q = .70$. To infer their genotype, we can expand the binomial term and plug in our empirical observations

$$1 = p^2 + 2pq + q^2$$

$$1 = .09 + .42 + .49$$

Punnett Square

One indispensible graphical tool of statistical investigation is the Punnett square. We can take our equations from above and create a Punnett square in the following manner

	p	**q**
p	pp	qp
q	pq	qq

The idea is to account for every possible combination of variables *p* and *q*, and assign a probability to them.

	p	**q**
p	.09	.42
Q	.42	.49

Q1: why does the diagonal axis (cells *pq* and *qp*) have the same probabilites associated with them? 3 points.

Q2: how would the values in the Punnett square change if the frequency of $p = .40$ and $q = .60$? Show your work*. 2 points.

From Structure to Function

Our probabilities from the 2x2 Punnett square represent the probability that any one offspring in the F_2 generation will inherit a particular genotype. However, not every cell in our matrix represents a distinct phenotype. Recall that our bird population has two phenotypes (red and blue) even though there are four possible genotypes. This can be possible because the *p* allele is dominant and the *q* allele is recessive. The rules of gene action in this case are the following:

* any genotype in which *p* exists results in a red phenotype.

* only genotypes in which no *p* allele exists can produce a blue phenotype.

Q3: According to these rules, list a) the genotypes that result in a red phenotype, b) the genotypes that result in a blue phenotype, and c) the total frequency at which each phenotype will occur. 4 points.

Multiple Loci and Epistasis: Agouti Mouse Example (16 points)

The Agouti mouse is a forest rodent that lives in Central and South America. They are a model organism for the effects of nutrition (via epigenetic effects) and epistatic effects on offspring coat color. This exercise will focus on the interactions of genes (epistasis) in determining the F_2 phenotype.

Now that we are familiar with a simple one locus model, we can move to model that includes two sites and two alleles. In the F_1 generation, there are two sites: one that confers coat color (*B/b*), and another (*C/c*) that determines albinism. The *B* allele is dominant and encodes for a black coat, while the *b* allele is recessive and encodes for a brown coat. The *C* allele allows for the expression of the coat color determined by *B/b*, while the *c* allele is **heterozygous recessive** and in such case can block the expression of the *B/b* site. This is shown below in graphical form:

B/b site

	B	**b**
B	BB (Black)	bB (Black)
b	Bb (Black)	bb (Brown)

C/c site

	C	**c**
C	CC (Agouti)	cC (Agouti)
c	Cc (Agouti)	cc (Albino)

Taken together, this results in a larger Punnett square and a more complex set of phenotypic relationships

	BB	Bb	bB	bb
CC	BBCC	BbCC	bBCC	bbCC
Cc	BBCc	BbCc	bBCc	bbCc
cC	BBcC	BbcC	bBcC	bbcC
cc	BBcc	Bbcc	bBcc	bbcc

Q1: List all possible genotypes for the following phenotypes: a) black coat, b) brown coat, c) albino. Calculate the expected frequency of each phenotype for F_2 (HINT: assume that each cell in the 4x4 Punnett square has a frequency of .0625). 4 points.

Q2: What are the "rules" of gene action (make a list like I did in the From Structure to Function section)? What criterion did you use to arrive at this classification (what results in a black coat, in a brown coat, etc.)? 4 points. Try to formulate as few rules as possible.

Q3: Suppose we have two specific Agouti mice: one with a black coat, and the other an albino. Further suppose that we know the genotype for each mouse: Black coat = BBCC, and Albino = BBcc. Tell me what the frequency of the following phenotypes in F2 will be for a) Black-coated offspring, b) Brown-coated offspring, c) Albino offspring. 4 points. HINT: the rules of gene action and independent assortment still apply.

Q4: Now assume that there are two parents with known genotypes: Black Coat = BBCC and Brown Coat = BbCC. Tell me what the frequency of the following phenotypes in F2 will be for a) Black-coated offspring, b) Brown-coated offspring, c) Albino offspring. 4 points.