

FALL 2023
Evolution in the Age of Genomics (GU4560)

Syllabus

STAFF Instructors: Guy Sella
 Fairchild 602A
 gs2747@columbia.edu
 Office hours: 2:30-3:30 pm Mon and additional times by
 appointment; please email beforehand.

Peter Andolfatto
 NWC 1313
 pa2543@columbia.edu
 Office hours: 2:30-3:30 pm Wed and additional times by
 appointment; please email beforehand.

 TA: Naima Okami
 NWC 1301
 neo2118@columbia.edu
 Office hours: 10:15-11:15am Wed and additional times by
 appointment; please email beforehand.

Lectures: Monday and Wednesday 1:10-2:25, Fairchild 1000

Sections: Wednesday 9:00-9:50, Fairchild 1000

COURSE DESCRIPTION: All life on earth has emerged from the evolutionary process, subject to a common set of rules. Principles of evolutionary genetics therefore provide the unifying theory that binds together what happens at the scale of individuals and populations, “microevolution,” with what plays out over paleontological timescales, “macroevolution”. These principles are relied on explicitly or implicitly across biological research, from studies of biodiversity to those of embryonic development, when the goal is to identify disease risk factors in humans or to predict the dynamics of new Sars-cov-2 variants. A solid foundation in evolution is thus fundamental to modern biology, whether in academic research, medicine, public health, anthropology or environmental policy.

This course is designed to provide that foundation, introducing basic concepts in Evolutionary Biology from speciation to natural selection. While the lectures incorporate a historical perspective, the main goal of the class is to familiarize students with topics and tools of evolutionary genetics as practiced today, in the age of genomics. Thus, the focus will

be on evidence from molecular evolution and genetics and exercises will assume a basic background in genetics. Examples will be drawn from across the tree of life, but with a primary focus on humans.

DRAFT COURSE OUTLINE (subject to change):

WEEK 1 (Sept 6 and 11)

GUY SELLA, Charles Darwin's "Origin of Species": These lectures explore why Darwin's work was so important and corrects some pervasive misconceptions about the theory of evolution. Along the way, we articulate the principle of natural selection; review important chapters in Darwin's life and work; ask ourselves what the principle of natural selection really explains on its own and what is it that the study of evolution aims to explain; discuss biology before Darwin; briefly review the pillars of Darwin's arguments in the Origin and explain why they were so radical and important; review the structure and content of the Origin; go over some misconceptions, such as: evolution equals optimization and that the notion of evolution is somehow still in question; review the two main problems that Darwin encountered with his theory (the age of the earth and the mode of inheritance).

WEEK 2 (Sept 13 and 18)

GUY SELLA, Evidence for evolution by descent and phylogenetic trees: We review evidence for several aspects of the evolutionary theory. Species change by means of natural selection: Peter and Rosemary Grant's work on Darwin's finches. New species are created in evolution: the ring species of salamanders in California and continuity between micro and macro-evolution. Species go extinct: Trilobites, Dinosaurs and "just so stories". Missing links: the Archaeopteryx and the search for Tiktaalik. Evidence for evolution on larger time scales: the principle of Uniformitarianism and the formation of the Grand Canyon. Evidence for common descent in the deep past: analogies and homologies, the structure of the hand in tetrapods, legs in snakes and whales, and the top 10 rudimentary organs. Evidence for a universal common ancestor of all life on earth: the standard genetic code and Crick's frozen accident. Correlated homologies: constructing phylogenetic trees from DNA data and the work of Penny et al. (1982). The UPGMA algorithm for phylogenetic reconstruction and its limitations. Reconstructing ancestral DNA sequences using maximum parsimony. The Sears tower of evidence for evolution in the genomic era and the tree of life.

WEEK 3 (Sept. 20 & 25)

PETER ANDOLFATTO, The fossil record & biogeography

Fossils, how they form and how we date them. The fossil record in (very) broad strokes. Five lessons from the fossil record, and popular misconceptions. Gradualism, saltationism and punctuated equilibrium. Biogeography. Island biogeography. Vicariance biogeography. Phylogeography. The case of marsupials. The great American interchange. Our own interchanges.

WEEK 4 (Sept 27)

GUY SELLA, Why population genetics? This lecture addresses the important role of mathematical models in evolutionary biology (all the math we will use is high school level). Evolutionary processes occur in populations across generations, and population genetics focuses on how the genetic composition of populations changes across generations. Darwin's problem of blending inheritance and early Mendelian genetics. The debate between early Mendelians and Biometricians, and R. A. Fisher's 1918 resolution. Fisher, Wright and Haldane, the modern synthesis, and the acceptance of evolutionary theory in biology. Why did mathematical models play a crucial role in the acceptance of evolutionary theory? Problems in direct measurement and observation; complex dynamic processes. The evolutionary forces that we will explore in subsequent lectures. Basic definitions (loci and alleles). How models of population genetics can be tested: the great obsession with polymorphism and divergence. Example: the ABO blood groups in humans and apes.

WEEKS 4-5 (Oct 2 & Oct 4)

GUY SELLA, Random genetic drift and mutation: The simplest possible model – one autosomal locus, two alleles, and an infinite panmictic population of constant size – predicts no change in allele frequencies. The Hardy-Weinberg relationships can be tested empirically, and used to detect selection and population structure. The Wright-Fisher process in finite populations: random walks and the elimination of genetic variation by Random Genetic Drift (RGD). RGD backwards in time: Common Ancestor, MRCA, and genealogy at a locus. The relationship between the genealogy at a locus and pedigrees, and why genealogies vary across loci. The mitochondrial Eve and Y-chromosome Adam: evidence for the African origin of modern humans. The relationship between heterozygosity, mutation rate, and population

size, and applications to conservation biology. The effective population size and other basic numbers (mutation rate, divergence, polymorphism) in humans, chimpanzees and *Drosophila*. Why is the effective population size in humans so small? Divergence under RGD and mutation: neutral fixation and the molecular clock. Different kind of mutations and their rates: point mutations, indels, microsatellites; different kind of changes are used to estimate times on different evolutionary time scales. Simulating RGD using the program POPULUS. Measuring diversity and divergence in DNA sequences, and using divergence to date phylogenetic trees. Testing the molecular clock using the relative rate test.

WEEK 5 (Oct 9)

PETER ANDOLFATTO, Human evolution and prehistory: What is unique about humans? Many types of humans. The pitfalls, challenges and history of studying human evolution. Darwin's descent of man. Closing in on humans: primates today and their evolution; apes today and their evolution. Resolving the relationship with our closest relatives: morphological methods, immunological molecular methods, cytological methods, LINES and SINES-based methods. Nucleotide sequence-based methods and incomplete lineage sorting.

WEEK 6 (Oct 11 & 16)

PETER ANDOLFATTO, Evolution on the human lineage: Phenotypic trends from physical anthropology and archeology: *Homo erectus*; archaic humans; Neanderthals; anatomically modern humans. The multi-regional and Out of Africa hypotheses. How many hominins? *Homo floresiensis*. Learning about human history from genetic variation: the effects of demographic processes on neutral genetic variation; models of changing population size and population structure. Extant patterns of variation in human populations (why are a million people in Africa more variable genetically than a billion in China). Fst and population structure in humans. Ancestry, populations and races. Inferring the geographic ancestry of an individual based on DNA sequences (Novembre et al. 2009). Evidence for admixture with Neanderthals and other archaic humans. The theme of admixture; what is a population?

WEEK 7 (Oct 18 & 23)

PETER ANDOLFATTO, Natural selection, part I: Dobzhansky's statement and the bloodless ice fish of Bouvet Island. The problem of observing adaptation in progress. Industrial melanism: Kettlewell's experiments and Haldane's calculations of the selective advantage.

The basic model of viability selection: change in allele frequency in one generation; deducing the selective advantage from changes in frequency; fixed points, directional, divergent and balancing selection and estimating selection coefficients under balancing selection.

WEEK 8 (Oct 25 and 30)

GUY SELLA, Natural selection, part II: We explore two parallel themes: we relax some of the previous simplifying assumptions of the model to study mutation-selection balance and selection in finite populations, and we suggest several resolutions to the apparent paradox presented by genetic diseases. The sickle cell trait and malaria: Anthony Alison and the first example of selection in humans. Mutation-selection balance: dominant (sickle cell) and recessive (spinal muscular atrophy). Alison's studies. The geographic relationship between sickle cell and malaria today. The life cycle of *Plasmodium falciparum* and the advantage of sicklers. Several origins of the sickle cell trait. G6PD and the Duffy protein. Malaria and human history. Resistance to chloroquine and DDT. The fixation probability of a beneficial allele (the "truck effect"). Time to fixation. Adaptation from new mutation and from standing genetic variation.

WEEK 9 (Nov 1 & 8) * Note: No class on Nov 6th**

GUY SELLA, Molecular Evolution: I) Purifying selection: The discovery of Archea, the three domains system, and the relationship between domains (S16 based trees of life). 500 conserved proteins over the three domains of life. The numbers of genes in different taxa. Conserved proteins between human and puffer fish and human and chimpanzees. Back of the envelope calculations about the number of substitutions expected under neutrality in the 500 conserved genes. The importance and prevalence of purifying selection. The conservation of amino acids alongside the prevalence of silent changes. Quantifying the intensity of purifying selection on DNA sequences. Back of the envelope calculation for the rate of pseudogenization. II) The neutral theory of molecular evolution: Synonymous and non-synonymous changes between humans and chimpanzees. Why is it reasonable to assume that the overwhelming majority of mutations are neutral and deleterious: improving a Lamborghini using a baseball bat. The neutralist interpretation of the ratio of non-synonymous to synonymous changes between humans and chimpanzees. Non-synonymous and synonymous changes in the influenza virus. Revisiting the molecular clock. Testing the neutral theory: McDonald-Kreitman (1991). Example: an elongation factor. III) Adaptation:

Example: Prm-2, a spermatid-associated gene (Wykoff et al. 2000). Detecting and quantifying adaptive evolution using the entries of the McDonald-Kreitman table. Adaptive evolution in different taxa (Eyre-Walker 2006). Estimating the fraction of functional non-coding DNA and testing whether most adaptive evolution occurs in coding or non-coding regions of the genome (Andolfatto 2005).

WEEK 10 (Nov 13)

PETER ANDOLFATTO, Gene duplications

Orthologs and paralogs. Origins of new genes. Mechanisms of gene duplications. Copy number variants are deleterious yet many duplicates are retained. Adaptive theories for the maintenance of gene duplicates (SRGAP2, antifreeze proteins, conotoxins). Non-adaptive theories: the duplication-degeneration-complementation model (HOX genes and a recent genomic evaluation). Balancing selection and gene duplication.

WEEK 11 (Nov 15 and 20)

GUY SELLA, How do new capabilities arise in evolution: the case study of color vision:

Color vision in different taxa and the evolutionary determinants of the number of genes. Why is color vision a good case study? The physics, biochemistry, anatomy and neurobiology of color vision in a nutshell. Visual pigments and opsins. Variation in the number of opsins in different taxa. Rephrasing the question: how are opsins gained? How do they change? Do new opsins confer a selective advantage? How and why are opsins lost? The evolution of tri-chromatic vision in Old World Monkeys: parsimony suggests a single origin; the genetic architecture shows a single duplication event; concordance between the molecular dating of duplication and the geological record of the continent separation. Evidence for selection: reduced amino acid divergence indicates purifying selection on the new opsin; amino acid substitutions indicate positive selection for frequency attenuation; scarcity of color blindness in the wild and behavioral studies also indicate selection. The origin of tri-chromatic color vision in howler monkeys: parsimony suggests a separate origin; the genetic architecture shows a different duplication event; concordance between molecular dating of duplication and the split from spider monkeys; amino acid substitutions and frequency attenuation show parallel evolution in OWM and howler monkeys. Parallel evolution as evidence for adaptation. The amazing story of tri-chromatic vision in the polymorphic females of marmosets and squirrel monkeys: X-inactivation explains how a single opsin is expressed in

each photoreceptor; the age of the alleles indicate that the polymorphism is under balancing selection; behavioral studies show that foraging troops take advantage of the tri-chromatic females. Color blindness and the polymorphic tri-chromatic females illustrate that the plasticity of the nervous system allows it to cope with changes in the number of opsins. Opsin death and parallel evolution in the deep sea. Revisiting the question about the number of genes in different taxa. Back of the envelope calculations for different gene duplication scenarios: neo-functionalization, the masking effect and sub-functionalization.

***** No class on Nov 22nd *****

WEEK 12 (Nov 27 and 29).

PETER ANDOLFATTO, Convergent evolution & Evo-devo

Convergence and parallelism at different scales, from phenotype to nucleotide substitutions. The “predictability” of evolution. How convergent evolution reveals constraints on evolution: mutation target size, epistasis and pleiotropy. Examples of the molecular basis for convergent evolution. The evolution of development. Deep homology. The evolution of novelty. The argument for the importance of regulatory evolution.

WEEK 13 (Dec 4 & 6)

PETER ANDOLFATTO, Speciation: The Biological Species Concept. Bacterial species. Polyploidy speciation in plants. Geographic modes of speciation. The Dobzhansky-Muller model of hybrid incompatibilities. The “snowball” effect (Orr 1995). The mapping of incompatibilities (Presgraves 2003). Speciation genes. Haldane’s rule and Muller’s dominance theory. The genic view of speciation.

WEEK 14 (Dec 11)

GUY SELLA and PETER ANDOLFATTO, Re-cap and Review.

READINGS will draw from research articles, popular science books and articles, and *The Origin of Species*. Either one research article or 30-40 pages of a book or popular article will be assigned per lecture. Some of the readings are given below:

Week 1: Chap. 14, *The Origin of Species* by Charles Darwin

Week 2: Chap. 10, *The Greatest Show on Earth: The Evidence for Evolution* by Richard Dawkins; Preface, *Making of the Fittest: DNA and the ultimate forensic record of evolution* by S. B. Carroll

Week 3: pp. 87-101 from Chap. 4: *The Origin of Life, Evolution* by Nicholas H. Barton; Chap. 1, *Finding Your Inner Fish* by Neil Shubin

Week 4: Chap. 1 & 2, *Making of the Fittest* by Sean Carroll

Week 5: Chap. 12 & 13 from Barton

Week 6: Chap. 1, *Human Variation*; "Sleeping with the Enemy" by Elizabeth Kolbert for *The New Yorker* (2011)

Week 7: "Skin Deep" by N. G. Jablonski and G. Chaplin for *The Scientific American* (2003)

Week 8: Chap. 3, *Making of the Fittest*

Week 9: McDonald & Kreitman (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*.

Week 10: Chap. 4, *Making of the Fittest*; Lynch & Conery (2000). The evolutionary fate and consequences of duplicate genes.

Week 11: Chap. 5 & 6, *Making of the Fittest*

Week 12: Chap. 8, *Making of the Fittest*; Stern & Orgogozo (2009). Is genetic evolution predictable?

Week 13: Chap. 7: *The Origin of Species*, from *Why Evolution is True* by Jerry Coyne; Zhen et al. (2012). Parallel molecular evolution in an herbivore community.

Week 14: Schumer et al. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species.

COURSE EVALUATION

The coursework will include weekly or bi-weekly exercises and readings, as well as a take home exam. The reading will be verified by short online quizzes before lectures. In order to complete the course, students will be required to submit at least 7 of 10 exercises, and pass 7 of 11 quizzes as well as the final exam. The grade of the course will be structured as follows:

- Up to 15% based on the grades of the quizzes (only the best seven grades will be included).
- Up to 35% based on the grades of the exercises (only the best seven grades will be included).
- Up to 40% based on the grade on the final exam (there will be no choices in the exam but the total number of points will be greater than 100). The exam will include questions and exercises from all the material of the course.
- Up to 10% based on class participation.