Evolution of Biological Molecules (HGEN/ECEV/BCMB 31100)

Winter 2018; Tu/Th 11:00-12:50
GCIS W105

Instructors
Joe Thornton, Professor, Departments of Human Genetics and Ecology & Evolution (joet1@uchicago.edu)
D. Allan Drummond, Assistant Professor, Department of Biochemistry & Molecular Biology (dadrummond@uchicago.edu)

Teaching Assistant
Yeonwoo Park, GGSB
(yeonwoo@uchicago.edu)

Course Description

This introductory graduate-level course connects evolutionary changes in genes and genomes with the structure, function and behavior of the encoded protein and RNA molecules. Central themes are the mechanisms and dynamics by which molecular structure and function evolve, how protein/RNA architecture shapes evolutionary trajectories, and how patterns in present-day sequences can be interpreted to reveal the interplay data of evolutionary history and molecular properties. Core concepts in macromolecule biochemistry (folding and stability of proteins and RNA, structure-function relationships, kinetics, catalysis) and molecular evolution (selection, mutation, drift, epistasis, effective population size, phylogenetics) will be taught, and the interplay between them explored. Students will derive and simulate evolution using molecular-level models; analyze empirical structure, function and sequence data in an evolutionary framework; and, through discussion and reading of classic and recent literature, trace the development of key ideas in molecular evolution up to the present day. The final project will involve application of course material to a relevant research question chosen by each student.

Course Objectives

Our goals are 1) to provide a serious introduction to evolutionary concepts for the biochemist and molecular biologist, 2) to provide a serious introduction to biochemical concepts for the evolutionary biologist, and 3) introduce both groups of students to an emerging paradigm of research at the interface of the two disciplines.

After this course, students will:

- have basic professional literacy in fundamental methods and concepts in population genetics, phylogenetics, molecular evolution, and protein biochemistry sufficient to comprehend scientific papers and pursue further study in these fields;
- know how to simulate evolutionary processes and protein folding under simple models;
- derive the statistical mechanical partition function, define the free energy of folding, and describe evolutionary constraints on the folded and unfolded ensembles; describe mutational effects in terms of Michaelis-Menten kinetics.
- define fitness and derive and utilize simple one-locus models of mutation, selection, and drift;
- align protein/RNA sequences, infer evolutionary changes using phylogenetic methods, map changes to

protein/RNA structures, and predict how those changes might affect protein structure, stability, substrate binding, and activity;

- define and utilize the concept of likelihood and probabilistic evolutionary models as applied to analysis of phylogenies, ancestral sequences, and testing models of sequence evolution.

Prerequisites

Comfort with basic computer programming (course will use Python and R); undergraduate biology, chemistry, calculus, and introductory statistics.

Format

The way to gain advanced understanding in a scientific field is to wrestle directly with the field's seminal concepts and their implementation in case studies from the literature. The course will consist of interactive lectures, discussions of primary readings from the literature, and problem sets. How productive and interesting the class will be is determined in significant part by how much everybody participates.

You are primarily responsible for your own learning: you have to work through the readings and problems until you really understand them, ask questions in class about anything you find challenging, and engage in discussions about the meaning, importance, and validity of each paper.

There are many issues that will only be covered in class, so your regular attendance is essential and required. This is not the kind of class in which you can learn or succeed by staying home and reading the book, because there is no book. (Not yet anyways.)

No electronic devices

In an effort to create a more focused and interactive learning environment, use of laptops and tablets will not be permitted during the lectures. We understand that such devices can be useful for note-taking, but they inevitably detract from the discussion-based environment we seek to create. Thus, we ask you to put away your devices during class, to take your notes on paper, and to bring printed copies of the papers assigned to each session.

Readings

There is no textbook on the evolution of biological molecules. The following three books may provide useful general background and supplement:

- Page RDM and Holmes EC. Molecular Evolution: A Phylogenetic Approach. Sinauer, 2009.
- Lesk AM, Introduction to Protein Science (2nd Edition). Oxford University Press, 2010.
- Haddock S and Dunn C. Practical Computing for Biologists, Sinauer, 2011.

Virtually all of the course readings will be from the primary literature. We will work through key points from the papers in class, so you will have a chance to bring up any issues that are not clear to you, as well as points that you think are worth discussing or challenging. Readings must be completed by the time of the class meeting in which they will be discussed. Class participation is essential to an intellectually lively course; it will also contribute

significantly to your grade.

Presentations

Each student will be required to host the discussion of one paper during the quarter. The host is responsible for leading the discussion and presentation of a paper in an interactive **chalk talk.** We believe that this format will be more rewarding and engaging for all students than the typical electronic presentation.

The host must be prepared to communicate the important aspects of the paper -- including its purpose, strategy, experimental/analytical methods, results, and conclusions -- using only notes and drawings to be made on the classroom whiteboard during the discussion. It will be impossible to recreate every figure from a paper in its original form. Rather, the host's goal should be to identify the essential aspects of each analysis and to display them in a simplified graphic format that can form the basis for a group discussion and assessment. If you are unsure as to why a particular paper is selected for discussion, and you wish to confirm that you are picking up on the proper thematic content, please reach out to the TA prior to your discussion section.

The presenter alone is not solely responsible for the discussion. Everyone should participate in the conversation to reach mutual understanding of the paper and to assess its contribution to the field. The professors may also spontaneously ask other students in the class to explain a result or analysis, so please be prepared to do so.

Problem sets

This is an upper-level course. The way to gain advanced understanding in a scientific field is to wrestle directly with the field's seminal concepts and their implementation in case studies from the literature. The course will consist of interactive lectures, discussions of primary readings from the literature, and problem sets. How productive and interesting the class will be is determined in significant part by how much everybody participates.

You are primarily responsible for your own learning: you have to work through the readings and problems until you really understand them, ask questions in class about anything you couldn't figure out, and engage in discussions about the meaning, importance, and validity of each paper.

Software

The course will use R and Python, along with a number of special tools. Some of these tools are available on web servers, and others will require installation on your computers. We will provide the URLs for obtaining the software, as well as links to manuals and tutorials. If you need help with installation, let us know. Everyone will need to install RStudio, PyMol, and Python. Details for installation of these programs is provided in the "Programming and Installation" document on Canvas.

Exam

There will be one take-home exam—a midterm—consisting of short essays and problems for you to work and explain. This is an open-book exam but, unlike the problem sets, it is absolutely <u>not</u> collaborative. You are to work on this exam <u>entirely on your own</u>, without any discussion with other students. There is no final exam.

Paper

You will write one in-depth term paper on a subject of your choice in the field of evolution of biological molecules. The proposal must explicitly apply approaches from both molecular biology/biochemistry and from molecular evolution. Preliminary analyses will involve applying methods introduced during the course, specifically:

phylogenetic analysis, ancestral reconstruction, mapping evolutionary variability to molecular structure, tests of selection, population genetics, analysis of evolutionary rates and/or codon usage.

The complete paper will be fully referenced and should be 12–20 pages in length (double-spaced). The paper will be in the form of a grant proposal with preliminary data and analyses. It should have the following structure (with very rough page guidelines provided):

1) **Specific aims/summary** (~1.5 – 2 pages, double-spaced). A summary of the importance and overall goal of the project and the specific aims that will be pursued and the hypotheses that will be tested. (There are typically 2-4 such aims in a project; each can have subparts if necessary.)

2) **Background and justification** (~2 – 4 pages). The purpose of this section is to justify your project via an explanation of its importance to the relevant scientific fields and a critical evaluation of the published literature on your topic. The goal is to explain clearly why the question is important and the model system being studied appropriate, what is known about it, how we know it, what questions are currently unresolved, and why the approach you propose will contribute to our knowledge.

3) **Preliminary data/analysis** (~2-3 pages). This section should include detailed description of the methods you have used to make preliminary progress towards your aims and testing your hypotheses. This may include phylogenetic, molecular evolutionary, structural modeling, simulation-based, or even experimental analyses. (Because this is a ten-week course, we recognize that most of you won't be able to conduct extensive bench work for the project.)

4) **Proposed research** (the rest! ~2-4 pages per aim, depending on the intricacy and number of aims). You should clearly state the questions you seek to answer, describe and justify your general strategy, and describe the specific kinds of data you would gather and what techniques you will use to analyze it. Most important is how you will interpret the data. Be sure to explicitly explain what you will conclude given the different possible outcomes of your analysis. What you propose should be practical, given the kinds of tools that you have learned about in the course. If you have not done so in the preceding sections, be sure to justify the choices you make about the tools and types of analyses you will use.

5) **Figures and references**—not included in page limit.

At the end of week 5, you must turn in at least one substantial idea for your proposal, summarized in a short ~3-sentence excerpt. If you would prefer to turn in multiple ideas and receive feedback from the instructors and TAs, you are free to do so. As always, we are all open for consultation at any point in the process of generating themes for your project.

A one-page prospectus is due by Tuesday of week 7. It should describe the problem you will be addressing and provide a brief justification of its importance.

The final paper is due to the instructors and TA by 5pm on March 17th (the last day of finals week).

This paper an exercise in scientific thinking and in mastering a subtopic relevant to the course. The grade will be based on 1) how clearly and persuasively you argue for the importance of your topic in its scholarly context, 2) the clarity and testability of your hypotheses, 3) the suitability of the analyses you propose for testing these hypotheses and the rigor of your discussion of how you will interpret the data, and 4) the design and rigor of your preliminary analysis and interpretation of the data it produces.

The clarity of your writing and the organization of your ideas is crucial. Scientists use the written word—in grant proposals and manuscripts—to convince each other of the value of their work. As graduate students, honing your

skills in this area should be a top priority.

Paper workshops

One of the best ways to improve one's scientific thinking and writing is to read others' works-in-progress and comment on them; the feedback helps the author, too, of course. In week 8, we will assign members of the class into workshop groups of 4 students each. You will meet in week 8 with your group to discuss your Specific Aims page. You will also meet by Thursday of week 10 in these groups to review entire drafts of each others' proposals. Everyone will read the papers of the other members of the group in advance and bring comments about what you liked and constructive suggestions for how it might be improved, structured through a provided form.

In addition to sharing your comments with your group, you must email your written reviews to the course TA by March 9th.

Evaluation and grades

This is a pass-fail course. At the end of the term, we will communicate to the registrar whether your work was sufficient to have earned credit in the course. Passing requires work and engagement to the standard of excellence expected of a graduate student at the University of Chicago. Although we do not provide a letter grade, we will provide extensive evaluation and feedback on your assignments, your presentations and participation.

The pass-fail policy is consistent with the requirements and policies of all relevant graduate programs of which we are aware. If you are concerned about this, please consult the administrator of your program, and we are happy to provide further information if necessary.

We do not give letter grades because we consider them residuals of a high-school format, where students are too often trained to work for grades as if they were the objective. We assume that you, as graduate students, are taking the course because it will enhance your development as a scientist and because you find the material of interest; that, along with your commitment to engage productively with the community of your peers and professors, should be sufficient motivation to do high-quality work. In the years since we established our pass-fail policy, we have been inspired by our students' work and engagement for its own sake.

Academic and scientific integrity

You are graduate students, so it should be unnecessary for us to include this section. We have a zero-tolerance policy for breaches of academic integrity. University rules on academic integrity will be strictly enforced. Please refamiliarize yourself with the definitions of cheating and plagiarism. If you have doubts or questions about acceptable academic/scientific conduct in any situation, we encourage you to discuss the situation with us in advance.

COURSE SCHEDULE

The scheduled lectures, corresponding readings and their value to understanding biochemistry and evolution, and notable dates for the course are outlined below. The papers that will be presented by students appear in bold.

1A: Course overview and protein space

Evolution concepts: Evolutionary conservation and homology; from organismal features to genomic features to individual molecular features; drift and selection; evolution as tinkerer, not engineer
Biochemistry concepts: Sequence space; scarcity of biologically useful sequences; Connection between mutability of molecule in the lab and evolutionary sequence diversity.
Readings:
Smith, J. M. (1970). Natural selection and the concept of a protein space. *Nature*
Jacob, F. (1977). Evolution and Tinkering. *Science*
Supplemental: Borges, J.L. (1941) The Library of Babel

1B: Protein functions and fitness

Evolution concepts: Evolution as a stochastic process; exponential growth; fitness as a latent value/propensity
Biochemistry concepts: Molecular underpinnings of fitness; antibiotic/pesticide resistance mechanisms; fitness measurements in the laboratory vs. evolutionary scales.
Readings:
1. Walkiewicz, Katarzyna, Andres S. Benitez Cardenas, Christine Sun, Colin Bacorn, Gerda Saxer, and Yousif Shamoo. 2012. "Small Changes in Enzyme Function Can Lead to Surprisingly Large Fitness Effects during Adaptive Evolution of Antibiotic Resistance." ***Proceedings of the National Academy of Sciences*** **109 (52): 21408–13.**
2. Harms, Michael J., and Joseph W. Thornton. 2013. "Evolutionary Biochemistry: Revealing the Historical and Physical Causes of Protein Properties." *Nature Reviews Genetics* 14 (8): 559–71.

2A: Fitness and kinetics

Evolution concepts: Fundamentals of selection; simulations of evolution
Biochemistry concepts: Michaelis-Menten kinetics
Readings:
1. Course write-up
2. Echave, Julian and Claus O. Wilke. 2017. Biophysical models of protein evolution: Understanding the patterns of evolutionary sequence divergence. *Annual Review of Biophysics* 46 (in preprint); http://biorxiv.org/content/early/2016/08/30/072223
3. Gunawardena, Jeremy. 2012. "Some Lessons about Models from Michaelis and Menten." *Molecular Biology of the Cell* 23 (4): 517.
Supplementary readings:
1. Orr, H. Allen. 2009. "Fitness and Its Role in Evolutionary Genetics." *Nature Reviews Genetics* 10 (8): 531–39.
2. Sella, Guy, and Aaron E. Hirsh. 2005. "The Application of Statistical Physics to Evolutionary Biology." *Proceedings of the National Academy of Sciences of the United States of America* 102 (27): 9541–46.

2B: Genetic drift, the neutral theory, and molecular clocks

Evolution concepts: basis and predictions of the neutral theory, functional density vs selective constraint
Biochemistry concepts: relationships between sequence changes and functional changes, functional

vs. selective constraints

Readings:

1. **Hietpas, Ryan T., Jeffrey D. Jensen, and Daniel NA Bolon. 2011. "Experimental Illumination of a Fitness Landscape."** ***Proceedings of the National Academy of Sciences*** **108 (19): 7896–7901.**

2. Kimura, Motoo 1986. "DNA and the Neutral Theory." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 312 (1154): 343–54.

3. King, Jack L., Thomas Hughes Jukes, and Bryan Clarke. 1969. *Non-Darwinian Evolution*. Bobbs-Merrill.

Supplementary readings:

1. Axe, Douglas D., Nicholas W. Foster, and Alan R. Fersht. 1996. "Active Barnase Variants with Completely Random Hydrophobic Cores." *Proceedings of the National Academy of Sciences* 93 (11): 5590–94.

2. Nei, Masatoshi, Yoshiyuki Suzuki, and Masafumi Nozawa. 2010. "The Neutral Theory of Molecular Evolution in the Genomic Era." *Annual Review of Genomics and Human Genetics* 11: 265–89.

3A: Genetic drift, the neutral theory, and molecular clocks

Evolution concepts: Evidence for and against molecular clocks

Biochemistry concepts: Biochemical underpinnings and assumptions of molecular clocks

Readings:

1. Lang, Gregory I., Daniel P. Rice, Mark J. Hickman, Erica Sodergren, George M. Weinstock, David Botstein, and Michael M. Desai. 2013. "Pervasive Genetic Hitchhiking and Clonal Interference in Forty Evolving Yeast Populations." ***Nature*** **500 (7464):571–74.**

2. Dickerson, Richard E. 1971. "The Structure of Cytochrome C and the Rates of Molecular Evolution." *Journal of Molecular Evolution* 1 (1): 26–45.

Supplementary readings:

1. Zuckerkandl, Emile. 1976. "Evolutionary Processes and Evolutionary Noise at the Molecular Level." *Journal of Molecular Evolution* 7 (4): 269–311.

2. Chothia, Cyrus, and Arthur M. Lesk. 1986. "The Relation between the Divergence of Sequence and Structure in Proteins." *The EMBO Journal* 5 (4): 823.

3. Bromham, Lindell, and David Penny. 2003. "The Modern Molecular Clock." *Nature Reviews Genetics* 4 (3): 216–24.

4. Rodriguez-Trelles, Francisco, Rosa Tarrio, and Francisco J. Ayala. "Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH."*Proceedings of the National Academy of Sciences* 98.20 (2001): 11405-11410.

3B: Phylogenetics I

Evolution concepts: Trees, likelihood, substitution models

Biochemistry concepts: Structure and function of reconstructed ancestral proteins

Readings: None.

4A: Protein folding, stability, and exchangeability

Evolution concepts: Principles underlying substitution matrices

Biochemistry concepts: Partition functions and quantitative models of folding and stability; driving forces underlying protein folding

Readings:

1. Jacquier, Hervé, André Birgy, Hervé Le Nagard, Yves Mechulam, Emmanuelle Schmitt, Jérémy Glodt, Beatrice Bercot, et al. 2013. "Capturing the Mutational Landscape of the Beta-Lactamase TEM-1." ***Proceedings of the National Academy of Sciences of the United States of America*** **110 (32):13067–72.**

2. Course write-up

3. Dayhoff, Margaret O., and Robert M. Schwartz. 1978. "A Model of Evolutionary Change in Proteins." In *In Atlas of*

Protein Sequence and Structure. Citeseer.
4. Bloom, J. D., J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, and F. H. Arnold. 2005. "Thermodynamic Prediction of Protein Neutrality." *Proceedings of the National Academy of Sciences of the United States of America* 102:606–11.

4B: Phylogenetics II

Evolution concepts: Ancestral reconstruction
Biochemistry concepts: Structure and function of reconstructed ancestral proteins redux
Readings:
1. Ortlund, Eric A., Jamie T. Bridgham, Matthew R. Redinbo, and Joseph W. Thornton. 2007. "Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis." ***Science*** **317 (5844): 1544–48.**
2. Harms, Michael J., and Joseph W. Thornton. 2010. "Analyzing Protein Structure and Function Using Ancestral Gene Reconstruction." *Current Opinion in Structural Biology* 20 (3): 360–66.

** Week 5: Topics due for final project

5A: Mutation-selection balance

Evolution concepts: Mutation-selection-drift balance; genetic load, optimality of evolution, and efficacy of selection
Biochemistry concepts: Exploration of why proteins are marginally stable and why transcription factors bind with less than maximum affinity
Readings:
1. **Taverna, Darin M., and Richard A. Goldstein. 2002. "Why Are Proteins Marginally Stable?"** ***Proteins: Structure, Function, and Bioinformatics*** **46 (1): 105–9.**
2. Sailer, Zachary R., and Michael J. Harms. 2017. "Molecular Ensembles Make Evolution Unpredictable." *Proceedings of the National Academy of Sciences of the United States of America* 114 (45):11938–43.
Supplementary readings:
1. Vy Nguyen, Christopher Wilson, Marc Hoemberger, John B. Stiller, Roman V. Agafonov, Steffen Kutter, Justin English, Douglas L. Theobald, and Dorothee Kern. 2016. "Evolutionary drivers of thermoadaptation in enzyme catalysis." *Science* 10.1126/science.aah3717
2. Shoichet, Brian K., Walter A. Baase, Ryota Kuroki, and Brian W. Matthews. 1995. "A Relationship between Protein Stability and Protein Function." *Proceedings of the National Academy of Sciences* 92 (2): 452–56.
3. Serrano, L., A. G. Day, and A. R. Fersht. 1993. "Step-Wise Mutation of Barnase to Binase. A Procedure for Engineering Increased Stability of Proteins and an Experimental Analysis of the Evolution of Protein Stability." *Journal of Molecular Biology* 233 (2):305–12.

5B: Protein folding and structural variability

Evolution concepts: Rates of evolution at surface vs. core; mutational effects on stability
Biochemistry concepts: Driving forces in protein folding; the hydrophobic effect
Readings:
1. Alexander, Patrick A., Yanan He, Yihong Chen, John Orban, and Philip N. Bryan. 2009. "A Minimal Sequence Code for Switching Protein Structure and Function." ***Proceedings of the National Academy of Sciences*** **106 (50): 21149–54.**
2. Worth, Catherine L., Sungsam Gong, and Tom L. Blundell. 2009. "Structural and Functional Constraints in the

Evolution of Protein Families." *Nature Reviews Molecular Cell Biology* 10 (10): 709–20.
Supplementary readings:
1. Pace, C. Nick, Bret A. Shirley, M. McNutt, and K. Gajiwala. 1996. "Forces Contributing to the Conformational Stability of Proteins." *The FASEB Journal* 10 (1): 75–83.
2. Pace, C. Nick, and J. Martin Scholtz. 1998. "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins." *Biophysical Journal* 75 (1): 422–27.
3. Scherrer, Michael P., Austin G. Meyer, and Claus O. Wilke. 2012. "Modeling Coding-Sequence Evolution within the Context of Residue Solvent Accessibility." *BMC Evolutionary Biology* 12 (September): 179.

6A: Functional constraints on rates of evolution

Evolution concepts: Purifying selection and neutrality; functional and structural constraints as a determinant of evolutionary rates; effect of dispensability
Biochemistry concepts: Empirical constraints on different structural domains
Readings:
1. Jack, Benjamin R., Austin G. Meyer, Julian Echave, and Claus O. Wilke. 2016. "Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes." *PLOS Biology* 14 (5):e1002452.

6B: "Non-functional" constraints on rates of evolution

Evolution concepts: Factors affecting rates of evolution beyond immediate function
Biochemistry concepts: Costs of protein production: ATP, time, toxicity of misfolding, designability
Readings:
1. Drummond, D. Allan, Jesse D. Bloom, Christoph Adami, Claus O. Wilke, and Frances H. Arnold. 2005. "Why Highly Expressed Proteins Evolve Slowly." *Proceedings of the National Academy of Sciences of the United States of America* 102 (40): 14338–43.
2. Levy ED, De S, Teichmann SA. 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proceedings of the National Academy of Sciences of the United States of America* 109(50):20461-6.

** Week 7: One-page specific aims due

7A: Epistasis I

Evolution concepts: Different forms of epistasis; implications for evolvability and robustness
Biochemistry concepts: Biophysical causes of epistasis
Readings:
1. **Gong, Lizhi Ian, Marc A. Suchard, and Jesse D. Bloom. 2013. "Stability-Mediated Epistasis Constrains the Evolution of an Influenza Protein." *Elife* 2.**
2. Harms, Michael J., and Joseph W. Thornton. "Historical Contingency and Its Biophysical Basis in Glucocorticoid Receptor Evolution." *Nature* 512, no. 7513 (2014): 203–7.
3. Olson CA, Wu NC, Sun R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. Curr Biol. 2014 24(22):2643-51.

Supplemental
Starr TN, Thornton JW. Epistasis in protein evolution. Protein Sci. 2016 25:1204, 2016.

7B: Epistasis II

Evolution concepts: Methods for finding patterns of covariation; phylogenetic non-independence
Biochemistry concepts: Structural and statistical coupling in protein structure; prediction of structure from patterns of covariation
Readings:
1. Skerker, Jeffrey M., Barrett S. Perchuk, Albert Siryaporn, Emma A. Lubin, Orr Ashenberg, Mark Goulian, and Michael T. Laub. 2008. "Rewiring the Specificity of Two-Component Signal Transduction Systems." Cell 133 (6): 1043–54.
2. Hopf TA, Ingraham JB, Poelwijk FJ, Scharfe CPI, SPringer M, Sander C, Marks DS. Mutation effects predicted from sequence co-variation. Nature Biotechnol. In press.
3. Ashenberg, Orr, L. Ian Gong, and Jesse D. Bloom. 2013. "Mutational Effects on Stability Are Largely Conserved during Protein Evolution." Proceedings of the National Academy of Sciences 110 (52): 21071–76.
Supplementary Reading:
1. Podgornaia, Anna I., and Michael T. Laub. "Pervasive degeneracy and epistasis in a protein-protein interface." *Science* 347.6222 (2015): 673-677.
2. Marks, Debora S., Thomas A. Hopf, and Chris Sander. 2012. "Protein structure prediction from sequence variation." (Perspective) *Nature Biotechnology* 30, 1072–1080.

8A: Adaptation

Evolution concepts: Statistical tests of selection; linkage and hitchhiking; distribution of mutational effect sizes
Biochemistry concepts: Empirical rationale for tests of selection; connecting signatures of selection and biochemical evolution
Readings:
1. **Barber, Matthew F., and Nels C. Elde. 2014. "Escape from Bacterial Iron Piracy through Rapid Evolution of Transferrin." Science 346 (6215): 1362–66.**
2. Starr TN, Picton L, Thornton JW. 2017. "Alternative Evolutionary Histories in the Sequence Space of an Ancient Protein" Nature 549:409-413.
3. Yokoyama, Shozo, Takashi Tada, Huan Zhang, and Lyle Britt. 2008. "Elucidation of Phenotypic Adaptations: Molecular Analyses of Dim-Light Vision Proteins in Vertebrates." Proceedings of the National Academy of Sciences 105 (36): 13480–85.
Supplementary Reading:
1. Fay, Justin C. 2011. "Weighing the Evidence for Adaptation at the Molecular Level." Trends in Genetics 27 (9): 343–49.
2. Bielawski, Joseph P., and Ziheng Yang. 2003. "Maximum Likelihood Methods for Detecting Adaptive Evolution after Gene Duplication." In Genome Evolution, 201–12. Springer.

8B: Gene duplication

Evolution concepts: Evolution of protein families; models for functional evolution after gene duplication. Why are duplicated genes retained so often? Where do new functions come from—and how do we know when functions are new?
Biochemistry concepts: Comparative studies of structure-function among orthologs and paralogs
Readings:
1. **Van Hoof, Ambro. 2005. "Conserved Functions of Yeast Genes Support the Duplication, Degeneration and Complementation Model for Gene Duplication." *Genetics* 171 (4): 1455–61.**
2. Hittinger, Chris Todd, and Sean B. Carroll. 2007. "Gene Duplication and the Adaptive Evolution of a Classic Genetic

Switch." *Nature* 449 (7163): 677–81.

** Week 8: Peer review sessions on specific aims

9A: Regulation

Evolution concepts: Conservation of regulatory sites; motif-finding
Biochemistry concepts: Regulation of protein activity by modification, allostery, and localization
Readings:
1. **Lynch, Vincent J., Gemma May, and Günter P. Wagner. 2011. "Regulatory Evolution through Divergence of a Phosphoswitch in the Transcription Factor CEBPB."** ***Nature*** **480 (7377): 383–86.**
2. Pearlman, Samuel M., Zach Serber, and James E. Ferrell Jr. 2011. "A Mechanism for the Evolution of Phosphorylation Sites." *Cell* 147 (4): 934–46.
Supplementary readings:
1. Holt, Liam J., Brian B. Tuch, Judit Villén, Alexander D. Johnson, Steven P. Gygi, and David O. Morgan. 2009. "Global Analysis of Cdk1 Substrate Phosphorylation Sites Provides Insights into Evolution." Science 325 (5948): 1682–86.

9B: Ligand interactions / Evolution of molecular complexes

Evolution concepts: Evolution of promiscuity and specificity; mechanisms for evolving new specificity and complexity
Biochemistry concepts: Characteristics of binding interfaces; effects of kinetics and thermodynamics on binding
Readings:
1. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR. 2013. "Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution." Proc Natl Acad Sci U S A. May 28;110(22):9007-12.
2. Eick, Geeta N., Jennifer K. Colucci, Michael J. Harms, Eric A. Ortlund, and Joseph W. Thornton. 2012. "Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors." *PLoS Genetics* 8 (11): e1003072.
Supplementary readings:
1. Fersht AR, Basis of Biological Specificity. Trends in Biochem Sci 9:145-147, 1984.
2. Siddiq MA, Hochberg GK, Thornton JW. Evolution of protein specificity: insights from ancestral protein reconstruction. Curr Opin Struct Biol. 2017 47:113-122.

10A: Disordered sequences and the formation of large protein assemblies

Evolution concepts: Signatures of selection in unalignable/"low-complexity" sequences
Biochemistry concepts: Disordered regions, protein self-assembly, phase separation
Readings:
1. Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, Sosnick TR, Drummond DA. 2017. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. Cell 168 (6) :1028–1040.
2. Frey, Steffen, Ralf P. Richter, and Dirk Görlich. 2007. FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. *Science* 314:815–817.

Supplemental readings:
1. Kato and others. 2012. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form

Dynamic Fibers within Hydrogels. *Cell* 149:743–57.