

# THE ORIGINS OF CELLULAR ARCHITECTURE

## INTRODUCTION

### 1. Evolutionary Cell Biology.

- The dominance of unicellular life.
- What is evolutionary cell biology?
- The completeness of evolutionary theory.
- Nonadaptive hypotheses and our understanding of evolution.
- The grand challenges.
  - The origin of life.
  - The roots of organismal complexity.
  - Molecular stochasticity.
  - Molecular complexes.
  - Cellular networks.
  - Cellular surveillance systems.
  - Growth regulation.
  - Biological scaling laws.

### 2. The Origin of Cells.

- The earliest stages.
  - The alkaline hydrothermal-vent hypothesis.
  - The terrestrial geothermal-field hypothesis.
- An early RNA world?
- Membranes and the origin of individuality.
- Genomic constraints on the establishment of life.

### 3. The Major Lines of Descent.

- The major domains of life.
- Times of origin.
- The emergence of eukaryotes.
  - The stem eukaryote.
  - The eukaryotic radiation.
  - A eukaryotic big bang?

## THE GENETIC MECHANISMS OF EVOLUTION

### 4. The Population-genetic Environment.

- Demystifying random genetic drift.
- The genetic effective sizes of populations.
- Probability of fixation of a mutant allele.
- Evolution of the mutation rate.
  - High mutability of mutation rates.

- Error-prone polymerases.
- Optimizing the mutation rate.
- The nonrandom nature of mutations.

Recombination.

- Evolution of the recombination rate.

### **5. Evolution as a Population-genetic Process.**

- The perils of the adaptive paradigm.
- The fitness effects of new mutations.
- The classical model of sequential fixation.
- Vaulting barriers to more complex adaptations.
  - Sequential fixation vs. stochastic tunneling.
  - Two-locus transitions.
  - More complex scenarios.
  - The effects of recombination.
- The phylogenetic dispersion of mean phenotypes.
  - Two-state traits.
  - Multistate-traits and the drift-barrier hypothesis.

### **6. Evolution of Cellular Complexity.**

- Illusions of grandeur.
- Constructive neutral evolution.
  - Ribosomes.
- Evolution by gene duplication.
  - The masking effect.
  - Neofunctionalization.
  - Subfunctionalization.
  - Adaptive-conflict resolution.
- The case for subfunctionalization.
- The emergence of modular gene subfunctions.
- The passive origin of species via gene duplication.

## **BASIC CELLULAR FEATURES**

### **7. The Cellular Environment.**

- Molecular composition of cells.
  - Water.
  - Elemental composition.
  - Biomolecules.
- Numbers of molecules per cell.
- Passive transport of particles through the cytoplasm.
- Intermolecular encounter rates.
- Temperature-dependence of biological properties.
- Energy, carbon skeletons, and cell yield.

## **8. Evolutionary Scaling Laws in Cell Biology.**

Describing allometric relationships.

Scaling laws in cellular bioenergetics.

Metabolic rate.

Lifetime energy requirements of a cell.

The speed limit on cell-division rates.

The limits to natural selection imposed by the drift barrier.

Membrane bioenergetics and the prokaryote-eukaryote transition.

Energy production and the mitochondrion.

Cellular investment in ribosomes.

The mitochondrion as a driver of eukaryotic evolution.

## **9. Cell Growth and Division.**

Ribosomes and cell growth.

Models for cellular growth.

Control of cell size at maturity.

Molecular mechanisms of division-size determination.

Environmental determinants of cell size.

Scaling of intracellular features.

Phenotypic variation in cell size and division time.

Stochastic partitioning of cell contents at division.

Phenotypic variation and adaptation.

Environmental variation and the efficiency of selection.

Inheritance of environmental effects.

The adaptive value of phenotypic variation.

## **10. The Cell Life Cycle.**

The eukaryotic cell cycle.

Phylogenetic diversity.

Network complexity.

Mitosis.

Meiosis.

Origin and evolutionary modifications.

Rapid evolution of meiosis-associated proteins.

Recombination mechanisms.

Evolutionary consequences of sexual reproduction.

Mating types.

Mating-type determination.

Mating-type number.

Cell fusion.

Coevolution of pheromones and their receptors.

Sexual systems in unicellular and multicellular organisms.

Isogamy vs. anisogamy.

Sex ratio.

Sex chromosomes.

## **11. Cellular Senescence.**

- Physiological load.
- Error catastrophe.
- Cellular vs. population immortality.
- Molecular and cellular determinants.
- Evolution of senescence.
- Mutational meltdown.

## **STRUCTURAL EVOLUTION**

### **12. The Protein World.**

- The essential features of proteins.
  - Amino-acid composition.
  - Origin of amino acids.
- Protein folding and stability.
  - The rate of protein folding.
  - Stability of folding.
- Determinants of protein-sequence evolution.
  - Lessons from phylogenetic comparisons and experimental mutagenesis.
  - Expression level and the propensity for sequence change.
  - Mutation pressure and biased amino-acid usage.
  - Epistasis and compensatory mutation.
- A general model for protein evolution.

### **13. Multimerization.**

- The incidence and architectural features of multimers.
- Propensity to aggregate.
  - Theory of association.
  - The physical features of interfaces.
- Evolutionary considerations.
  - Transitions from monomeric to higher-order states.
  - The domain-swapping model.
  - Heteromers from homomers.

### **14. Protein Management.**

- Chaperone assistance.
  - Phylogenetic diversity of chaperones.
  - Client-chaperone coevolution.
  - Chaperone-mediated phenotypic evolution.
- Protein disposal.
- Post-translational modification.

### **15. Lipids and Membranes.**

- Molecular structure.
- Membrane structure.

Eukaryotes and the endogenous organellar explosion.

Vesicle trafficking.

    Vesicle production.

    Vesicle delivery.

    Evolutionary issues.

The nuclear envelope.

    Nuclear-pore architecture.

    Nuclear transport.

    Evolutionary considerations.

## **16. Cytoskeleton, Cell Shape, and Motility.**

The basic cytoskeletal infrastructure.

    Actins.

    Tubulins.

    Intermediate filaments.

Cell shape.

Cell walls.

Molecular motors.

Motility.

    Crawling.

    Prokaryotic flagella.

    Eukaryotic flagella.

    The costs of swimming.

## **ENERGETICS AND METABOLISM**

### **17. The Costs of Cellular Features.**

The bioenergetic cost of a cellular feature.

The evolutionary cost of a cellular feature.

An empirical shortcut to cost estimates.

The energetic cost of a gene.

    Chromosome-associated costs.

    Transcription-associated costs.

    Translation-associated costs.

    Evolutionary implications.

The cost of lipids and membranes.

    Cost of individual molecules.

    Total cellular investment.

### **18. Resource Acquisition and Homeostasis.**

Adaptive fine-tuning of elemental composition.

Nutrient uptake kinetics.

    Channels and transporters.

    Physiological acclimation.

    Advantages of motility.

Photosynthesis.

The transformation of solar to chemical energy.

The world's most abundant enzyme.

Osmoregulation.

Circadian rhythms.

## **19. Enzymes and Metabolic Pathways.**

Enzymes.

Basic enzymology

Degree of molecular perfection.

Enzyme promiscuity.

Pathway flux control.

Pathway position and the strength of selection.

Speed vs. efficiency.

Pathway expansion and contraction.

Stochastic meandering of pathway architecture.

The origin of novel enzymes.

Pathway participant remodeling.

Nonorthologous gene replacement.

Internal pathway expansion via multifunctional enzymes.

Pathway duplication and divergence.

## **INFORMATION PROCESSING**

### **20. Intracellular Errors.**

Transcript fidelity

Translational fidelity.

Biophysics of substrate discrimination and the cost of proofreading.

The limits to selection on error rates.

The evolutionary consequences surveillance-mechanism layering.

Adaptive significance of errors.

### **21. Intracellular Communication: Transcription.**

Transcription factors and their regulatory motifs.

Biophysics of recognition.

Facilitated diffusion and the search for regulatory motifs.

Evolution of the regulatory vocabulary.

Evolutionary rewiring of transcription networks.

### **22. Environmental Sensing and Extracellular Communication.**

Bacterial signal-transduction systems.

Origin and diversification.

Coevolutionary integration of components.

Emergence of new pathways.

Interconvertible proteins and ultrasensitivity.

The cost of signal transduction.  
Similarities and differences in eukaryotic systems.  
Chemotaxis.  
Accuracy of environmental assessment.  
Phenotypic bimodality and bet-hedging.  
Adaptive fine-tuning vs. inadvertent by-products of pathway structure.

## **ORGANISMAL COMPLEXITY**

### **23. Endosymbiosis.**

Mitochondria.  
Origins.  
Energetic boost or burden.  
Functional remodeling.  
The extreme population-genetic environments of mitochondria.  
Mutation rates.  
Modes of inheritance.  
Muller's ratchet.  
Organelle genome degradation.  
Animal mitochondrial tRNAs.  
Coevolutionary drive and compensatory mutations.  
Plastid evolution.  
Addiction to endosymbionts.

### **24. Origins of Organismal Complexity.**

Deconstructing the great chain of being.  
Genome complexity and organismal complexity.  
A shake-up of genomic organization in the ancestral eukaryote.  
Multicellularity.  
Multicellularity and cooperativity in bacteria.  
The costs of multicellularity.  
The emergence of cell-type specialization.  
Closing comments.

## FOUNDATIONS

- 2.1. The proton-motive force and the evolution of ATP synthase.
- 2.2. Evolution prior to self-replication.
- 2.3. The limits to replication fidelity and genomic maintenance.
  
- 4.1. The amount of neutral nucleotide variation maintained at selection-drift equilibrium.
- 4.2. Relationship of the recombination rate to physical distance between sites.
  
- 5.1. Divergence under uniform selection.
- 5.2. Mean probabilities of alternative alleles at steady state.
- 5.3. The detailed-balance solution for the evolutionary distribution of alternative molecular states.
  
- 7.1. Intracellular diffusion.
- 7.2. Rates of encounter by molecular diffusion.
- 7.3. The Boltzmann probability distribution for alternative molecular states.
- 7.4. The yield of cellular biomass per ATP usage.
  
- 8.1. The cost of building a cell.
- 8.2. Connecting metabolic rates with growth potential.
  
- 9.1. The scaling of ribosome number and cell growth rate.
- 9.2. Nutrient limitation and cell growth.
- 9.3. Scaling models for the development of cellular features.
- 9.4. Parent-offspring resemblance and the response to selection.
- 9.5. Transient response to selection without genetic change.
  
- 11.1. The physiological damage load in a cell lineage.
  
- 13.1. Association / dissociation equilibria.
  
- 14.1. The CCT chaperonin complex.
- 14.2. Evolution of a digital trait.
  
- 15.1. Probability of preservation and subdivision of labor by duplicated interactions.
  
- 16.1. The eukaryotic cellular investment in the cytoskeleton.
- 16.2. The physical challenges to cellular locomotion.
- 16.3. The construction costs of flagella.
  
- 17.1. The relationship of bioenergetic costs to the strength of selection.
- 17.2. The biosynthetic costs of nucleotides and amino acids.
- 17.3. The mutational hazard of excess DNA.
- 17.4. Numbers of molecules required in a cellular lifespan.
- 17.5. The biosynthetic costs of lipid molecules.

18.1. The response of uptake rate to nutrient concentration.

18.2. Encounter and capture rates.

18.3. The cost of osmoregulation.

19.1. Michaelis-Menten enzyme kinetics.

19.2. Evolutionary sensitivity of pathway steps.

19.3. Optimization of the glycolytic flux rate.

19.4. Extension / contraction of a metabolic pathway.

20.1. Kinetic proofreading.

20.2. The evolutionary bounds on the transcript-error rate.

21.1. Number of transcripts per cell.

21.2. Occupancy probability for a transcription-factor binding site.

21.3. The biophysics of TFBS localization.

21.4. The evolutionary dispersion of TFBS matching profiles.

22.1. Behavior of a monocycling system.

22.2. Accuracy of environmental sensing.

23.1. Messenger RNA editing.

## FIGURES

- 1.1. A broad overview of the Tree of Life.
- 1.2. The major dimensions of the triad of environmental features influencing cellular evolution.
  - 2.1. Amino-acid synthesis via the Strecker reaction.
  - 2.2. The reductive citric-acid cycle.
  - 2.3. Two proposed settings for the origin of life.
  - 2.4. The structure of ATP synthase.
  - 2.5. The building blocks of DNA and RNA.
  - 2.6. A family of polymeric sequences with binary alternatives for the monomeric subunits.
- 3.1. An idealized view of the two-domain view of the Tree of Life.
- 3.2. Alternative forms of phospholipids deployed by the three major lineages of life.
- 3.3. An approximate phylogenetic tree for some of the major eukaryotic supergroups.
- 3.4. Emergence of reproductive incompatibility following the relocation of a mitochondrial gene.
- 4.1. An overview of the influence of random genetic drift on allele frequencies.
- 4.2. Effects of chromosomal linkage on the efficiency of natural selection.
- 4.3. The negative scaling of effective population size with organism size across the Tree of Life.
- 4.4. The excess mutation load in an asexual mutator strain.
- 4.5. Quasi-equilibrium distributions of mutation rates under the drift-barrier hypothesis.
- 4.6. Negative scaling of mutation rates with effective population sizes across the Tree of Life.
- 4.7. The physical mechanics of recombination.
- 4.8. Scaling of the recombination rate per nucleotide site with genome size.
- 4.9. Temporal changes of allele frequencies under asexual vs. sexual reproduction.
- 5.1. The distribution of fitness effects of new mutations.
- 5.2. Allele-frequency changes in experimental populations of *E. coli*.
- 5.3. Some possible routes to the establishment of adaptations involving two or more mutations.
- 5.4. Origin of a complex adaptation involving three mutations in small and large populations.
- 5.5. Expected frequency of an allele under the joint forces of drift, mutation, and selection.
- 5.6. Schematic for transition rates between adjacent classes under the sequential-fixation model.
- 5.7. Long-term mean genotypic states under a two-locus, two-allele model.
- 5.8. Equilibrium mean frequencies of favorable alleles for a trait under directional selection.
- 6.1. Evolution of increased complexity by constructive neutral evolution.
- 6.2. Variation in the structural complexity and size of ribosomal RNA.
- 6.3. The DDC model for the alternative fates of duplicate genes.
- 6.4. Evolution of the yeast galactose-utilization pathway following duplication and subfunctionalization.
- 6.5. Duplication and subfunctionalization of components of yeast vacuolar ATP synthase.
- 6.6. The emergence of modular gene-regulatory structure by small duplications and deletions.
- 6.7. Speciation by reciprocal silencing of an ancestral duplicate gene.
- 7.1. Relationship between dry weights and volumes of individual cells.
- 7.2. Physical features of water molecules.

- 7.3. Fractional contributions of major biomass components to the cell dry weights.
- 7.4. Numbers of proteins and messenger RNAs per cell.
- 7.5. Two-dimensional random walks.
- 7.6. Diffusion coefficients for simple substances in water.
- 7.7. Response of cellular growth rates to temperature in bacterial species.
- 7.8. Heats of combustion of organic substrates and relationship to microbial growth rates.
- 7.9. Geometry of diffusive encounters between two particles.

- 8.1. Distribution of cell sizes for major phylogenetic groups.
- 8.2. Allometric scaling of metabolic rate with cell volume.
- 8.3. Estimating the costs of building and maintaining cells with chemostat cultures.
- 8.4. The bioenergetic costs of cellular construction and maintenance.
- 8.5. The scaling of maximum growth rate and cell volume in heterotrophs and autotrophs.
- 8.6. The drift-barrier to the performance of complex quantitative traits.
- 8.7. The scaling of mitochondrial features with cell size.
- 8.8. The scaling of ribosome number with cell size.
- 8.9. Allometric regressions of maximum growth rate and metabolic rate in ciliates.

- 9.1. Response of ribosome-associated features with cell growth rate.
- 9.2. Growth and uptake responses to the concentration of a limiting nutrient.
- 9.3. Relationship between cell volume and growth increment under three alternative models.
- 9.4. Evidence in support of the adder growth model for *E. coli*.
- 9.5. Alternative conceptual models for the determination of cell size at division.
- 9.6. Response of growth rate and cell volume to laboratory selection in *E. coli*.
- 9.7. Responses of cell size and growth rates at the phylogenetic and environmental plasticity levels.
- 9.8. Distributions of cell division times for different members of a population of *Bacillus*.
- 9.9. Coefficients of variation for numbers of molecules within individual cells.
- 9.10. Response to directional selection on a quantitative trait.
- 9.11. Influence of the form of the fitness function on mean fitness in a population.
- 9.12. A conceptualize partitioning of cellular proteins into functional sectors.

- 10.1. The eukaryotic cell cycle.
- 10.2. Nonorthologous gene replacement under a constant network topology.
- 10.3. Alternative surveillance architectures with the same final outcome.
- 10.4. Independent duplication / subfunctionalization events for a key cell checkpoint gene.
- 10.5. Mitosis vs. meiosis.
- 10.6. Some multimeric proteins involved in DNA replication.
- 10.7. The synaptonemal complex.
- 10.8. The centromere-drive hypothesis.
- 10.9. Evolution of mating-type determination in a pheromone-receptor system.
- 10.10. The binuclear genomes in ciliates.
- 10.11. Neutral drift of a one-to-one communication system.
- 10.12. Frequency-dependent selection for alternative mating types.

- 11.1. Serial buildup and dilution of cellular damage.
- 11.2. The distribution of damage inheritance in cell pedigrees.

- 11.3. Passive relocation of a protein aggregate to the poles of cells.
  - 11.4. Survivorship curves for two yeast species.
  - 11.5. Continuous decline of cell-division potential in single-cell lineages.
  - 11.6. History of division times in *E. coli* cells.
- 
- 12.1. Guide to the amino acids.
  - 12.2. Structure of polypeptide chains.
  - 12.3. Biosynthetic pathways of amino acids.
  - 12.4. Alpha helices and beta sheets.
  - 12.5. Dependence of folding rate on amino-acid chain length.
  - 12.6. Evolution of population mean phenotypes on a hyperbolic fitness function.
  - 12.7. Frequency distributions of the destabilizing effects of mutations on proteins.
  - 12.8. Bacterial phylogenetic distribution of folding stabilities of dihydrofolate reductase.
  - 12.9. Exchangeability of amino acids in  $\beta$ -lactamase.
  - 12.10. Distributions of fitness effects from single amino-acid substitutions.
  - 12.11. Genome-wide G+C compositions relative to the neutral expectation.
  - 12.12. A restricted path of evolution of a three-residue protein resulting from epistasis.
- 
- 13.1. Some examples of the varied forms of multimeric complexes.
  - 13.2. Structures of monomers, dimers, and tetramers for the case of homomeric proteins.
  - 13.3. Relative frequencies of the major classes of protein structures across the Tree of Life.
  - 13.4. Structural topologies of heteromers involving two subunits per repeat.
  - 13.5. Binding strengths associated with interfaces of heterodimeric complexes.
  - 13.6. Effectively neutral evolution of new binding configurations at an interface patch.
  - 13.7. An idealized model for evolution along a linear array of alternative quaternary states of a protein.
  - 13.8. The domain-swapping model in a diploid population.
  - 13.9. Evolution of heterodimers from a homodimeric state following gene duplication.
- 
- 14.1. An idealized cross-sectional view of protein processing by the GroEL chaperone.
  - 14.2. A simplified view of some of the challenges to the evolution of a heteromeric ring molecule.
  - 14.3. An idealized view of the recognition-sequence space for client proteins of a chaperone.
  - 14.4. The ubiquitin-proteasome pathway for protein degradation.
  - 14.5. Some of the N-end rules for the acquisition of protein-degradation signals.
  - 14.6. Schematic for the transition rates between adjacent classes under the sequential-fixation model.
- 
- 15.1. Structures of the two major classes of lipid molecules.
  - 15.2. Lipid bilayers.
  - 15.3. Stages in a vesicle-transport pathway.
  - 15.4. The organelle paralogy hypothesis.
  - 15.5. Higher-order structure of the coats of eukaryotic lipid vesicles.
  - 15.6. Evolutionary diversification of the known adaptor proteins and the COPI coat subunit.
  - 15.7. A hierarchical view of the nuclear envelope and the nuclear pore complex.
  - 15.8. Scaling of the number of nuclear pores with nuclear volume, and genome size with cell volume.
  - 15.9. Probabilities of the fates of two pairs of interacting duplicate genes.
- 
- 16.1. The three key forms of fibril-forming proteins.

- 16.2. A motor protein interfacing with a microtubule.
  - 16.3. Gross structure of bacterial and eukaryotic flagella.
  - 16.4. Scaling relationships between swimming velocities and cell volume in unicellular species.
  - 16.5. Schematic of a peptidoglycan layer.
- 
- 17.1. Distinction between direct and opportunity costs associated with building-block synthesis.
  - 17.2. Evolutionary distinction between construction/maintenance cost and direct benefits of a trait.
  - 17.3. Costs for cellular building blocks from biochemical pathways and degree of chemical reduction.
  - 17.4. Distribution of gene-specific costs of replication, transcription, and translation in four species.
  - 17.5. Costs (relative to total cellular energy budgets) for average genes in prokaryotes and eukaryotes.
  - 17.6. Scaling of surface areas of internal membranes with cell size in eukaryotes.
  - 17.7. Major cellular pathways for production of the three major carriers of energy and reducing power.
  - 17.8. Key steps and energy demands in the biosynthesis of ribonucleotides and deoxyribonucleotides.
- 
- 18.1. A structural explanation for the use of dual high/low-affinity transporters for nutrient uptake.
  - 18.2. The historical record of the Earth's atmospheric CO<sub>2</sub> and O<sub>2</sub> concentrations.
  - 18.3. Some compatible solutes deployed in osmotic regulation.
  - 18.4. The molecular network for the circadian clock in the cyanobacterium *Synechococcus*.
- 
- 19.1. Michaelis-Menten enzyme kinetics as a function of substrate concentration.
  - 19.2. Distributions of kinetic features of enzymes surveyed across the Tree of Life.
  - 19.3. The canonical pathway of glycolysis.
  - 19.4. Alternative pathway topologies for acquisition of a key metabolite.
  - 19.5. A hyperbolic fitness function as a function of resource availability.
  - 19.6. Long-term evolutionary probabilities of the three alternative pathway states.
  - 19.7. The primary pathways of amino-acid biosynthesis.
  - 19.8. Alternative biosynthetic pathways in different phylogenetic lineages for three amino acids.
- 
- 20.1. Functions that must be successfully navigated for the production of a proper protein.
  - 20.2. Reaction steps involving two competing reactions involving right and wrong substrates.
  - 20.3. Reaction dynamics for the situation in which there is proofreading of an intermediate complex.
  - 20.4. The distribution of transcript-error rates across the Tree of Life.
  - 20.5. Evolutionary consequences of the addition of a secondary line of defense.
- 
- 22.1. Bacterial one-component, two-component, and phospho-relay systems for signal transduction.
  - 22.2. Scaling of number of response regulator and histidine kinases in two-component systems.
  - 22.3. Requirements for the preservation and long-term divergence of a duplicated HK-RR pair.
  - 22.4. Generalized scheme for a monocyclic cascade.
  - 22.5. Response of the active fraction of an interconvertible protein to external-ligand concentration.
  - 22.6. Three variants on the structure of signal-transduction pathways commonly found in eukaryotes.
  - 22.7. Idealized schematic of the chemotaxis pathway in *E. coli*.
  - 22.8. Equilibrium activity levels of interconvertible enzymes.
- 
- 23.1. Idealized variants of classical mitochondria, and a few of their key modifications.
  - 23.2. Number of protein-coding genes in the genomes of organelles and related bacterial lineages.
  - 23.3. Muller's ratchet of the accumulation of deleterious mutations in an asexual population.

- 24.1. Scaling of maximum growth rate with size at maturity across the Tree of Life.
- 24.2. Scaling of the number of protein-coding genes per genome vs. total genome size

## **TABLES**

- 7.1. Elemental compositions of cells.
- 8.1. Geometric features of common cell shapes.
- 9.1. Coefficients of variation for growth-related features of cells.
- 12.1. Properties of amino acids.
- 15.1. Fractional contributions of lipid molecules to plasma membranes in select species.
- 17.1. Bioenergetic costs for the synthesis of lipid molecules.
- 17.2. Contributions of membranes to total cellular growth costs.
- 20.1. Estimated error rates associated with translation.

## **SUPPLEMENTAL TABLES**

4.1. Mutation-rate and effective population size estimates.

7.1. Dry weights and volumes of individual cells.

7.2. Fractional contributions of major biomass components to the cell dry weights.

7.3. DNA content vs. cell size.

8.1. Caloric content of cells.

8.2. Number of ribosomes per cell in different species.

15.1. Relationship between genome size and cell volume.