Evolutionary cell biology: Two origins, one objective

Michael Lyncha,1, Mark C. Fieldb,2, Holly V. Goodsonc,2, Harmit S. Malikd,e,2, José B. Pereira-Lealf,2, David S. Roosg,2, Aaron P. Turkewitzb,2, and Shelley Sazertf,1

aDepartment of Biology, Indiana University, Bloomington, IN 47405; bDivision of Biological Chemistry and Drug Discovery, University of Dundee, Dundee DD1 5EH, United Kingdom; cDepartment of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556; dDivision of Basic Sciences and eHoward Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98195; fInstituto Gulbenkian de Ciência, P-2781-901 Oeiras, Portugal; gDepartment of Chemical Biology, University of Pennsylvania, Philadelphia, PA 19143; hDepartment of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637; and iVerna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030

Edited by W. Fred Doolittle, Dalhousie University, Halifax, NS, Canada, and approved October 20, 2014 (received for review September 3, 2014)

All aspects of biological diversification ultimately trace to evolutionary modifications at the cellular level. This central role of cells frames the basic questions as to how cells work and how cells come to be the way they are. Although these two lines of inquiry lie respectively within the traditional domains of cell biology and evolutionary biology, the two fields are actually deeply connected. The origin of cells constituted one of life’s most important early evolutionary transitions, simultaneously enabling replicating entities to corral the fruits of their catalytic labor and providing a unit of inheritance necessary for further evolutionary refinement and diversification. The centrality of cellular features to all aspects of biology motivates the focus of cell biology on the biological/biochemical aspects of a broad swath of traits that include gene expression, metabolism, intracellular transport and communication, cell–cell interactions, locomotion, and growth. No one questions the rich contributions that have resulted from this focus on how cells work. However, with an emphasis on maximizing experimental consistency in a well-characterized model systems, cell biologists have generally eschewed the variation that motivates most questions in evolutionary biology.

Because all evolutionary change ultimately requires modifications at the cellular level, questioning and understanding how cellular features arise and diversify should be a central research venue in evolutionary biology. However, if there is one glaring gap in this field, it is the absence of widespread cell-biological thinking. Despite the surge of interest at the molecular, genomic, and developmental levels, much of today’s study of evolution is only moderately concerned with cellular features, perhaps due to lack of appreciation for their wide variation among taxa. However, a full mechanistic understanding of evolutionary processes will never be achieved without an elucidation of how cellular features become established and modified.

The time is ripe for bridging the gap between the historically disconnected fields of cell biology and evolutionary biology and integrating them with the principles of biophysics and biochemistry into a formal field of evolutionary cell biology. Recent advances in cell-biological analysis and the acquisition of omic-scale datasets have broadened the opportunities for research on nonstandard model organisms, thereby facilitating the incorporation of phylogenetic diversity into cell-level studies. Our vision for this synthesis is motivated by the growing realization in both communities that an intellectual merger will yield dramatic increases in our understanding of cell-biological structures, functions, and processes, as well as insights into the cellular basis for evolutionary change. Although not an exhaustive list, the following questions motivate and illustrate the potential for this new field.

Why Are Cells the Way They Are, and Why Aren’t They Perfect?

Although it is easy to marvel about the refined features of cells and their robustness to perturbations (1), the field of bioengineering imagines and implements more efficient cellular mechanisms in extant organisms. What, then, limits the levels of molecular/cellular refinements that have been achieved by natural selection?

To What Extent Is Cell Biology Beholden to Historical Contingency?

We have learned an enormous amount about the genetic mechanisms of evolution since Darwin, and it remains true that evolution is an opportunistic process of “descent with modification.”


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.


1To whom correspondence may be addressed. Email: mlynch@indiana.edu or ssazergbcmc.edu.
working with the resources made available in previous generations. Once established, useful features cannot be easily dismantled and reassembled de novo unless there is an intermediate period of redundancy.

One remarkable example of how history continues to influence today’s cell biology is the near universal use of ATP synthase as a mechanism for energy generation (2). Embedded in the surface membranes of bacteria and organelar membranes of eukaryotes, this complex molecular machine uses the potential energy of a proton gradient to generate a rotational force that converts ADP to ATP, much like a turbine converts the potential energy of a water gradient into electricity. However, the proton gradient does not come for free: cells first use energy derived from metabolism to pump protons out of membrane-bound compartments, creating the gradient necessary for reentry through ATP synthase. Even assuming that ATP production is an essential requirement for the origin of life, it is by no means clear that the path chosen for ADP-to-ATP conversion is the only possibility.

Rather, the universal reliance of all of life on this mechanism of energy conversion may be a historical relic of the exploitable energy source present at the time of life’s foundation: e.g., a precellular period in which energy acquisition derived from a natural proton gradient between overlying low-pH marine waters and the alkaline interiors of vent mounds (3, 4). Despite the central significance of ATP synthase to bioenergetics across the Tree of Life (8), Comparative biochemical and protein-structural analysis within a phylogenetic framework has great potential to address many outstanding questions in this area, including whether variation in the multimeric states of proteins is a simple consequence of stochastic mutations of adhesive interface residues, with minimal effects on catalytic efficiency.

Similar questions arise about the biophysical properties of supermolecular structures, such as microtubules, actin filaments, and the endomembrane systems of eukaryotic cells (9). The self-assembly of lipid bilayers emerges spontaneously from the biophysical properties of amphiphilic molecules, and recent origin-of-life research suggests that some of the key first steps in the origin of life, such as the assembly and division of vesicles, are inevitable consequences of the behavior of organic molecules in water (10, 11).

Finally, general biophysical phenomena are undoubtedly involved in the patterning of phenotypes at the whole-cell level. For example, constraints on surface/volume scaling may have been involved in the establishment of internal membranes and their above-noted associations with bioenergetics (12). Such constraints may also have played a central role in the evolution of cell size and features of the nuclear envelope (13). The emergence of the nuclear envelope may have, in turn, had secondary evolutionary consequences, such as the establishment of a permissive environment for intron proliferation (7), which requires efficient pretranslational splicing of transcripts.

Although the preceding observations suggest that the emergence and diversification of numerous cellular features may be predictable on biophysical grounds alone, the imposition of constraints on a complex trait need not preclude substantial opportunities for modifying the underlying components, as previously discussed with respect to ATP synthase. For example, although there are common organizational principles in diverse regulatory, signal-transduction, and metabolic pathways, dramatic cases of rewiring have been revealed with the expansion of molecular and cell biological investigations to multiple species. Such examples include aspects of mating-type specification (14, 15), meiosis (16), cell cycle (17, 18), biosynthetic pathways (19–23), protein transport (24), nuclear organization (25), and ribosome production (26, 27). These kinds of observations imply that there are often numerous degrees of freedom for reorganizing the underlying determinants of otherwise constant cellular processes.

How Much of Cellular Complexity Is the Result of Adaptation? A commonly held but incorrect stance is that essentially all of evolution is a simple consequence of natural selection. Leaving no room for doubt on the process, this narrow view leaves the impression that the only unknowns in evolutionary biology are the identities of the selective agents operating on specific traits. However, population-genetic models make clear that the power of natural selection to promote beneficial mutations and to remove deleterious mutations is strongly influenced by other factors. Most notable among these factors is random genetic drift, which imposes noise in the evolutionary process owing to the finite numbers of individuals and chromosome architecture. Such stochasticity leads to the drift-barrier hypothesis for the evolvable limits to molecular refinement (28, 29), which postulates that the degree to which natural selection can refine any adaptation is defined by the genetic effective population size. One of the most dramatic examples of this principle is the inverse relationship between levels of replication fidelity and the effective population sizes of species across the Tree of Life (30). Reduced effective population sizes also lead to the establishment of weakly harmful embellishments such as introns and mobile-element insertions (7). Thus, rather than genome complexity being driven by natural selection, many aspects of the former actually arise as a consequence of inefficient selection.

Indeed, many pathways to greater complexity do not confer a selective fitness advantage at all. For example, due to pervasive duplication of entire genes (7) and their regulatory regions (31) and the promiscuity of many proteins (32), genes commonly acquire multiple modular functions. Subsequent duplication of such genes can then lead to a situation in which each copy loses a complementary subfunction, channeling both down independent evolutionary paths (33). Such dynamics may be responsible for the numerous cases of rewiring of regulatory and metabolic networks noted in the previous section (34, 35). In addition, the effectively neutral acquisition of a protein–protein-binding interaction can facilitate the subsequent accumulation of mutational alterations of interface residues that would be harmful if exposed, thereby rendering...
what was previously a monomeric structure permanently and irreversibly heteromeric (8, 36–39). Finally, although it has long been assumed that selection virtually always accepts only mutations with immediate positive effects on fitness, it is now known that, in sufficiently large populations, trait modifications involving mutations with individually deleterious effects can become established in large populations when the small subset of maladapted individuals maintained by recurrent mutation acquire complementary secondary mutations that restore or even enhance fitness (40, 41).

One goal of evolutionary cell biology should be to determine whether these general principles involving effectively neutral paths of molecular evolution extend to even higher-order biological features, such as intracellular architecture (37). Is natural selection a sufficient or even a necessary explanation for the evolution of the complex features of the ribosome, the spliceosome, the nuclear-pore complex, and the Golgi apparatus? Or is a march toward increased, and potentially irreversible, cellular complexity an inevitable outcome of mutation pressure and the inefficiencies of selection processes in finite populations?

The points raised above are not meant to suggest that structures as complex as ribosomes or ATP synthase are maladaptive. Certainly, today’s cells cannot survive without such molecular machines. However, the existence of complex cellular features need not imply that each of the myriad of changes that sculpted such structures over evolutionary time was adaptive at the time of establishment. The determination of whether it is even feasible for a cellular innovation to have been promoted by purely adaptive processes cannot be made in the absence of information about the population-genetic environment: i.e., the magnitudes of the power of mutation, recombination, and random genetic drift. All three features vary by orders of magnitude across the Tree of Life and can only roughly be inferred for ancestral species. Uncertainty in this area is a major challenge for evolutionary cell biology (30, 42).

**How Do Cellular Innovations Arise?**

For practical reasons, cell biology has historically focused on the average features of the members of large populations of genetically uniform cells. However, natural selection does not operate directly on population means but on variation among individuals. Moreover, the evolutionary response to selection on a trait is not a simple matter of variation, but a function of the fraction of variation that has a genetic basis. Estimation of these key parameters is now within reach as new technologies allow assays of single cells in a high-throughput manner. Applications of these methods to genetically uniform populations reveal substantial cell-to-cell variation in gene-specific numbers of transcripts and proteins in all domains of life (43–45), and such variation (intrinsic cellular noise) seems to be a natural outcome of biophysical features of interactions between transcription factors and their binding sites, which can be quantified in mechanistic terms (46, 47). These kinds of observations, which can be extended to other intracellular traits (48), are essential to understanding the limits to the evolvability of cellular features. This is because environmental variance (intracellular noise) reduces the ability of a population to respond to selection by overshadowing the heritable genetic component of variation (49).

Although conceptually straightforward, resolving the degree to which variation (and covariation) of phenotypes in populations of cells is a consequence of genetic vs. environmental causes will require large-scale experimental designs including genetically variable isolates. When applied in this way, single-cell phenotyping down to the level of individual molecules has the potential to revolutionize the field of quantitative genomics by elucidating the precise sources of variation underlying the expression of higher-order cellular features. Notably, the statistical framework of quantitative genomics is also fully equipped to address the evolutionary consequences of transient epigenetic effects (49), whose influences are dissipated over time with various levels of reinforcement (e.g., refs. 50–52).

**Where Do Cellular Innovations Map onto the Tree of Life?**

A first step in nearly all studies in evolutionary biology is the elucidation of phylogenetic patterns of variation. Although a purely historical perspective cannot reveal the mechanisms by which evolution proceeds, it does clarify what needs to be explained. Traditional cell biology is largely devoid of comprehensive comparative analyses, but recent studies demonstrate the power of such approaches, as illustrated by the following three examples.

The first example addresses the evolutionary origins of the network of organelles and underlying molecular features by which membrane trafficking emerged in eukaryotes. The sorting of proteins and lipids among the intracellular compartments of eukaryotic cells is mediated in part by a family of protein complexes called adaptins. Although it had been accepted for over a decade that there are only four adaptin complexes in eukaryotes, comparative genomics suggested the presence of a fifth highly divergent adaptin-like complex across eukaryotes (53). Subsequent characterization of the protein in human cells identified its cellular location and function, thereby fundamentally altering our basic understanding of vesicle-transport systems and the likely order of evolutionary events leading to their diversification. An even more recent phylogenetic analysis suggests the existence of a sixth form of adaptin complex (54), raising the possibility that still more remain to be discovered, perhaps with some complexes being restricted to a subset of taxa. A second striking example of the power of comparative analysis to inform our basic understanding of cell biology involves the discovery of an evolutionary relationship between what were considered two very different kinds of membrane-deformation proteins. Cargo transport in eukaryotic cells involves the use of diverse pathways initiating with membrane-coated vesicles supported by clathrin, and the cage forming proteins of cytoplasmic coat protein complexes I and II (COPI and COPII). Although these proteins are lacking in amino acid sequence similarity, comparative structural analysis suggests a common molecular architecture that is also related to the membrane-curving proteins involved in both the nuclear-pore complex (NPC) (55) and the adaptins discussed above. The structural and functional insights emerging from these observations guided the development of a mechanistic understanding of the NPC (56) and yielded a novel evolutionary proposal—the “pro-tocoatomer” hypothesis, which postulates that many vesicle-coating complexes and the NPC arose by descent with modification (55). Among other things, this concept has provided a potential explanation for how the diverse body plans of eukaryotic cells could have arisen from a simpler prokaryote-like ancestor.

In a third example, an integration of molecular and morphological phylogenetic analysis has led to the identification of novel components of centrioles and cilia, as well as to evolutionary hypotheses for how their coordinated biogenesis and functions in different cellular contexts have been achieved through duplication and divergence of an ancestral gene set (57, 58).

This small set of examples illustrates the considerable potential for more elaborate comparative analyses to elucidate the evolutionary foundations of the most basic eukaryotic cellular features. Of course,
The ascertainment of where cell-biological innovations map onto the Tree of Life and inference of phylogenetic points of gain and loss of various modifications will require a substantial increase in taxonomic sampling of cellular diversity. Of the estimated 5–100 million extant species, only ~1.5 million have been described at even a rudimentary level, and most of these taxa are heavily biased toward plants, animals, fungi, and microbes with direct human impact (59) (Fig. 1). Future studies of biodiversity are likely to continue to extend to the discovery of novel phyla for quite some time (e.g., refs. 60–62). These issues, together with the fact that typically about a third of predicted protein-coding genes per sequenced genome are undefined and/or restricted to narrow taxonomic groupings, make clear that we are still missing immense swaths of information on cellular diversity. This “missing phylogeny” is likely of high value to applied research efforts in medicine, agriculture, and environmental science.

Unfortunately, parts lists inferred from genome information alone can take us only so far. Although results from transcriptomics, metabolomics, etc. can provide additional information, such work must ultimately be coupled to detailed studies of individual gene products in diverse taxa. To this end, we envision the need for a new grand challenge in biology, such as the proposed Atlas of the Biology of Cells (www.nsf.gov/publications/pub_summ.jsp?ods_key=biol2009). The fundamental idea here is to develop a database for cellular/subcellular features for a judiciously chosen, phylogenetically broad set of organisms, with the goal of sampling the functional diversity of metabolic and cellular morphological traits in the fullest possible sense. To be maximally productive, such an enterprise will require the further development of automated, generalizable, and high-throughput cell-biological methods. Significant support for appropriate phylogenetic sampling, development of reliable culture methods, and standardized measurement methodology will also be necessary. Most importantly, the latter will require the establishment of not only controlled vocabularies and ontologies to provide a conceptual framework for data comparison, but also quantitative metrics for defining, comparing, and predicting cell-biological structures and processes.

The payoffs of such an organized research program are likely to be substantial. As an analogy to where evolutionary cell biology is and where it might lead, consider that whole-genome sequencing was barely a dream 25 y ago but, in the past decade, has revolutionized virtually every aspect of biology, vastly increasing our understanding of human-genetic disorders, methods for disease control, energy production, and ecosystem function. Such advances continue to inspire the development of new “omics technologies with enormous increases in accuracy and efficiency, as well as the emergence of novel computational technologies for storage, integration, and analysis that facilitate the rapid transformation of data into knowledge.

How Can Effective Implementation of Lessons from Evolutionary Cell Biology Be Ensured?

Cell biology textbooks traditionally focus on structures and pathways perceived to be common to all cells, only occasionally addressing specializations in individual phylogenetic lineages, and even more rarely mentioning their modes of diversification. In effect, we have built up a sort of canonical molecular and cell biology based on a few serendipitously selected model organisms. How things work in Escherichia coli, Saccharomyces cerevisiae, Drosophila melanogaster, and mouse cells is all too often viewed as the “normal” mode of biology, with differences observed in other organisms often being viewed as little more than amusing oddities. Imagine what today’s biology might look like if our models had been Nanoarchaeum (archaeabacteria), Paramecium (ciliate), Ceratium (dinoflagellate), and Pinus (gymnosperm).

The view that intracellular structures are essentially invariant in diverse organisms engenders the false impression that an evolutionary biologist has little to gain by pursuing studies at the cellular level. Moreover, the few statements about evolution that can be found in cell-biology textbooks and journal articles frequently speculate on the adaptive significance of cellular features, oversimplifying and obscuring our understanding of evolutionary mechanisms (42, 63). This outmoded view of evolutionary processes still gives rise to major misunderstandings, with substantial implications (64).

In summary, we have attempted to highlight why bridging the conceptual gap between cell biology and evolutionary biology is likely to enrich our understanding of virtually all biological processes. For example, although the natural spatial delimitation of cell biology resides at the cell membrane, an understanding of the evolutionary roots of various cellular features is of central relevance.
to evolutionary developmental biologists concerned with the origin of cell types (65). Evolutionary cell biology has a particularly high potential for informing a variety of practical matters with ecological, economic, and health benefits. Such applications include the facilitation of drug development and the elucidation of the mechanisms of drug sensitivity and resistance, and of the identification of the mechanisms of nutrient fluxes through the environment and their dependence on species-specific features. The removal of real and perceived conceptual and communication barriers (including those engendered by the use of specialized vocabularies) and the design and implementation of cross-disciplinary educational initiatives are central keys to building an interactive community of scientists essential for igniting an effective field of evolutionary cell biology.

Acknowledgments. We thank W. Ford Doolittle for helpful comments. This paper was, in part, inspired by the helpful comments. This paper was, in part, inspired by the helpful comments. This paper was, in part, inspired by the helpful comments. This paper was, in part, inspired by the helpful comments. This paper was, in part, inspired by the...

[References and acknowledgments follow...]