Handbook of Digital Innovation. edited by Nambisan, Lyytinen, Yoo
Edward Elgar | 253–265

14. The bits and bytes of biology: digitalization fuels an emerging generative platform for biological innovation

Rob J. Kulathinal, Youngjin Yoo and Sudhir Kumar

The biggest innovations of the 21st century will be at the intersection of biology and technology. A new era is beginning. (Steve Jobs)

1. INTRODUCTION

Digital innovation has permeated many academic disciplines from economics and management science to the digital humanities and is quickly transforming traditionally retrospective fields of inquiry into dynamic, predictive, and highly translational landscapes. Digital technology facilitates the rapid development of products, processes, and services (Frankelius, 2009) by inventing new frameworks and recombining current practices (Boland, Lyytinen, & Yoo, 2007). The generative nature of digital innovation further serves as a catalyst for future discovery and opportunities for innovation (e.g. Um, Yoo, Wattal, Kulathinal, & Zhang, 2013). And as fields become more digitally integrated, it is expected that innovations will be generated at increasingly higher rates.

Perhaps, no field has been affected more by pervasive digital innovation than the biological sciences. The rich, nuanced, and increasingly exhaustive data that describe all aspects of life are interconnected on two axes: a diversity axis that connects all species (including those that went extinct) via evolution, and a complexity axis that connects hierarchical levels of biological organization in an organism to its DNA code via "omics." Biology is quickly becoming a completely integrated digital platform with each of its diverse and complex biological components mappable to each other. Indeed, the last few decades have witnessed unprecedented discoveries that have resulted from the generativity of a connected digital framework across all aspects of biology.

In this work, we claim that biology is on the verge of a major transformation in digitality (e.g. Negroponte, 1995) with both intended and unintended positive consequences on our ability to innovate. From data to analysis and from theory to prediction, the biological sciences are becoming fully digitized and we argue that this transformation is occurring due to the very nature of biology: the genotype to phenotype map that defines organismal complexity is a universal feature that itself is evolutionarily linked through a single common ancestor and its digitizable code. A fully integrated digitalized biology and its associated mathematization (Cartwright, Giannerini, & González, 2016) provides an innovation platform allowing researchers to easily move from pattern to process to prediction to product.

2. A TRADITION OF DOCUMENTING LIFE

Biology has historically been a descriptive science. Steeped within an empirical tradition of observation, early biologists strived to understand the world around them by documenting its grandeur. The Scientific Revolution was largely defined by advances in scientific knowledge gained through observation and experience (e.g. John Locke, *An Essay Concerning Human Understanding*, 1689). Early anatomists attempted to describe the physical mechanics and morphological nuances of our bodies in painstaking illustrative detail. Pioneering microbiologists such as Antonie van Leeuwenhoek used microscopes to describe never before seen microorganisms. Botanists such as Carl Linnaeus provided new categorizations of life's vast diversity through taxonomy (Linnaeus, 1735). Globe-trotting naturalists such as Alexander von Humboldt and Louis Agassiz as well as Charles Darwin and Alfred Wallace, co-founders of the theory of evolution by natural selection, described the diversity of life around us, making sense of the complex patterns of life found on our planet through careful and detailed data collection, often bringing samples back to museums for public consumption. Biologists of all types were driven by the excitement of discovering new patterns of described life.

However, a primary challenge evident during the early days of most descriptive sciences was the lack of interconnectivity between nodes of knowledge. As a consequence, accumulated data were largely siloed and, often, independently archived via print media in public and private libraries. In biology, knowledge was relayed through curated but sequestered collections found in museums and universities. Like other disciplines, it was not until the advent of the Internet that the interconnectivity of biology via its digital nature was able to be fully realized, globalized, and democratized.

Divided into a growing number of specialized domains, biology was in dire need of a unifying theoretical framework that linked different species and their characteristics (i.e. phenotypes). Knowledge garnered across countless biological systems remained, for the most part, largely disconnected from each other. For example, what value would a mammalian developmental biologist find in discussing morphogenesis in early embryos with an entomologist? Why would a botanist interested in cellular growth and division discuss regulatory mechanisms with a mycologist who solely studies growth patterns in fungi? How does cellular substructure in bacteria relate to that found in our own cells? Such divides were evidenced in the establishment of different societies, institutes, and epistemologies that followed the Scientific Revolution during the Age of the Enlightenment. This anachronistic legacy is still observed in many colleges, universities, museums, and other institutes of higher learning where separate departments such as zoology, botany, microbiology, and genetics coexist on the same college campus, and often still without much connection with each other. A paradigm-changing theory was sorely needed to comprehend the interconnectivity of life.

3. CONNECTING ALL OF LIFE THROUGH COMMON ANCESTRY

With the publication of *On the Origin of Species* (1859), Charles Darwin transformed the view of the world we live in by linking the diversity of life—both past and present—and making it clear that taxonomic divides were self-imposed constructs (Figure 14.1). The body plans of elephants and their cohabiting parasites share physiological commonalities that are hundreds

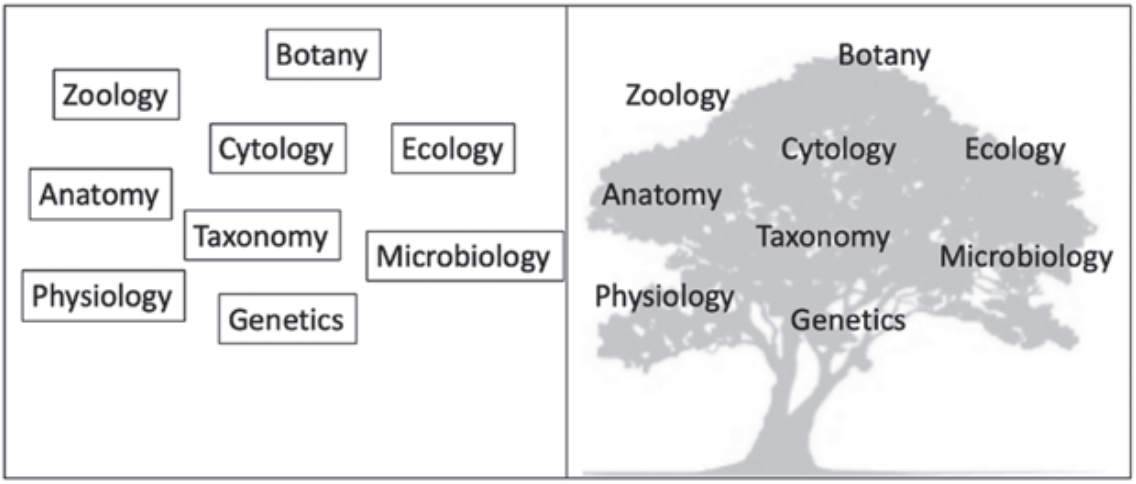


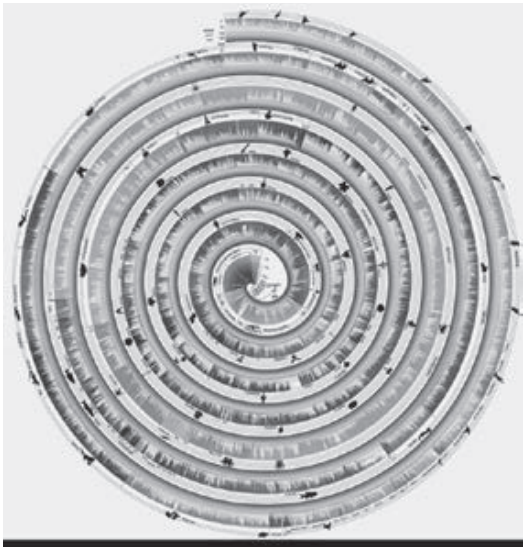
Figure 14.1 *Evolution transformed a set of independent and seemingly unrelated branches into the interconnected and integrated field of modern biology*

of millions of years old. Homologies in cellular processes between species as distinct as wheat and mushrooms can provide a comparative framework useful to each domain: cellular subcomponents may have been modified but are essentially the same in function. Darwin's grand theory forced biologists to consider the evolutionary processes underlying the patterns of organismal diversity based on their common ancestral history. Anatomical, physiological, and developmental mechanisms can now be interconnected and related to each other across phylogenetic space. The vertical transmission of traits between taxonomic families can be interpolated across species. Taxa as diverse as plants, animals, and bacteria are not only connected via homology (Figure 14.2) but are expected to share common molecular and evolutionary mechanisms.

Yet like many grand theories, it took years for scientists to fully appreciate the relevance, scope, and scale of the theory of evolution. In fact, it wasn't until the 1940s when several scientists from broad fields were able to finally break apart from the disciplines they identified with to assemble a "Modern Synthesis" (Dobzhansky, 1937; Huxley, 1942; Mayr, 1942; Simpson, 1944; Stebbins, 1950). Common patterns found between quite distinct species can now be seen to share common underlying processes of biology and phylogenetic inferences could be made about their origins and histories. Life on this planet as we know it is related and it is this very interrelatedness that makes all branches in biology—from zoology to botany to microbiology to genetics—connected to each other (Figure 14.1). As one of the Modern Synthesis founders famously wrote, "Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973).

4. MAPPING THE COMPLEXITY OF LIFE TO A GENETIC CODE

While the theory of evolution and the ensuing modern evolutionary synthesis explained the connections across life's diversity axis, the informational map that coded the genotype to phenotype complexity axis was still unknown. Until the middle of the last century, linking the complexity of life across hierarchical levels of biological organization within an organ-

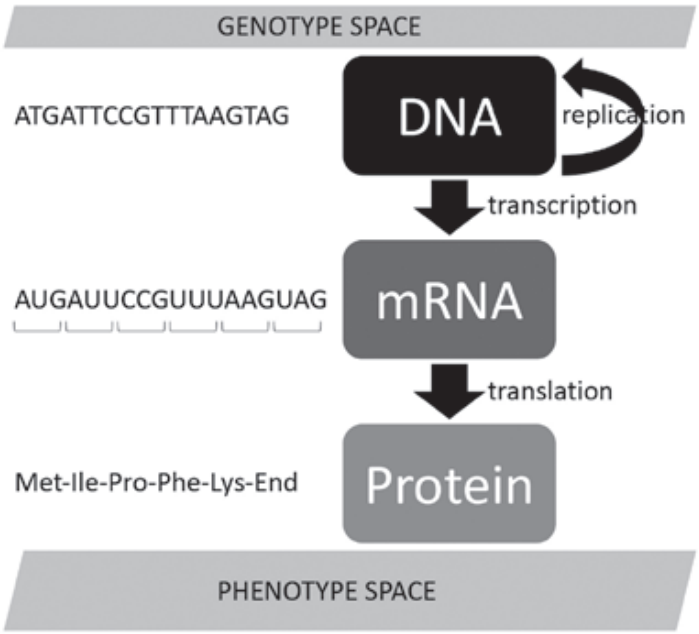


Note: The Tree of Life represents over 50,000 species demonstrating the interconnectivity of all life from a common ancestor over three billion years ago.
Source: Kumar, Stecher, Suleski, & Hedges (2017). A high-resolution image can be accessed at http://www.timetree.org.

Figure 14.2 Tree of Life

ism remained a major challenge. Specifically, it was not possible to map the complex set of character states—from DNA and proteins to cells, tissues, and organ systems—that define an individual organism. As a consequence, silos remained among scientists studying different levels of biological organization much like those found among scientists interested in different taxonomic groups. For example, early geneticists didn't talk to cell biologists who didn't interact with physiologists. Furthermore, it was not helpful that each taxonomic group and the specialized researchers who studied them applied their own taxon-specific dictionaries. This lack of a common ontological vocabulary across species meant that common processes could not be connected and comparatively assessed.

The discovery of DNA as the heritable unit of life in 1953 by James Watson, Francis Crick, and Rosalind Franklin, and the subsequent development of molecular biology during the latter half of the last century, finally provided us with a universal digital vocabulary of life based on the central dogma of molecular biology: DNA → mRNA → protein→ phenotype (Crick, 1970). During this time, it was found that the code of life provided a universal syntax that links the complexity found at these different hierarchical levels of organization (e.g. Nirenberg & Matthaei, 1961). Molecular biology heralded a new era of digitizable biology. A triplet codon drawing from only four bases could be combined in 64 (i.e. 4^3) ways providing a taxon-wide code that exceeds the 20+ available amino acids common to all life, from multicellular plant and animals to unicellular bacteria to the viruses that infect each of them. Linear combinations of these amino acids form proteins, the building blocks of life (Figure 14.3). So finally, the genetic code of life was cracked and easily digitized. And when combined, evolutionary theory and molecular biology's central dogma connect both biological axes—diversity and complexity—via a digitized code.



Note: From a linear array of nucleotides (A, T, C, G), 64 potential combinations of triplet codons code for ~21 amino acids, the building blocks of life and their characteristics (i.e. phenotypes).

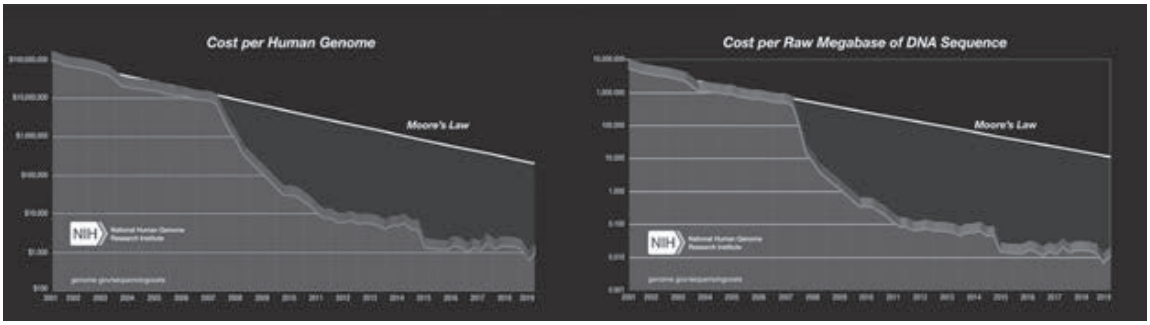
Figure 14.3 The central dogma of molecular biology

5. SYSTEMS BIOLOGY DRIVES DIGITALIZATION

Evolution and genetics, together, offered both a theoretical and methodological framework for the integrated study of life. However, it was not until very recently that we were able to fully exploit this framework. During the last two decades, the emergence of genomics and other "omic" technologies provided a powerful genome-wide extension of this framework that links all aspects of life on this planet—from complexity to diversity (e.g. Letunic & Bork, 2007; The International HapMap Consortium, 2003). This systems biology approach and in particular, integrated genomics, allows biologists to quickly and comprehensively map the unit of heredity across hierarchical levels of biological organization from DNA to global gene and protein expression to the phenotypic trait itself. Using nucleotide as well as amino acid sequences defined by a universal digital code, biologists can now connect co-expressing molecules involved in certain traits (Eisen, Spellman, Brown, & Botstein, 1998) and link gene networks to a diverse molecular ecosystem involved in regulating various functional components of an organism (Ashburner et al., 2000) via digitized character states.

Systems biology and the omics revolution has driven the digitization of biology in the following ways:

1. *Systems biology extends a genetic code that is finite and common*. Applying a standard library of four letters (A, G, T, C) that are evolutionary linked genome-wide enables omics to identify patterns and processes, both molecular and evolutionary, that can be extrapolated across all lineages. This combination of a shared common vocabulary across life and an evolutionary theory that connects our shared biology provides a powerful, simple, and

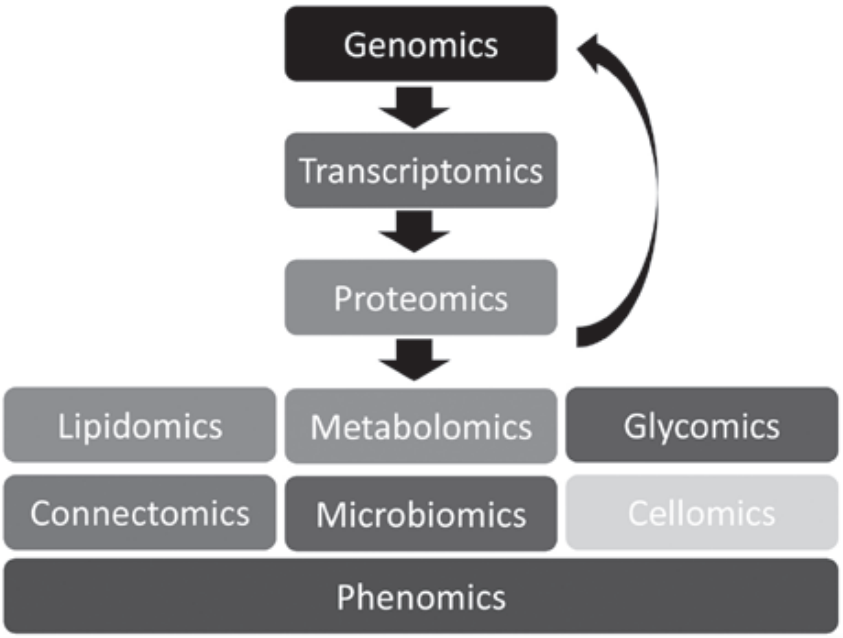


Source: Data taken from Wetterstrand (2019).

Figure 14.4 Decreasing costs of genomic sequencing powers the omics revolution

digital framework to study biology. Furthermore, the same standard set of assays, database schemas, analysis tools, etc. can be employed on all species lineages using this universal syntax.

2. *The democratization of "omics."* The development of new genomic sequencing technologies (Margulies et al., 2005, Shendure et al., 2005, Bentley et al., 2008) made possible the seamless transition in biology from a gene-centric to genomics-based framework. Today, generating genome assemblies and associated omics data is fast, efficient, and cost-effective, making biology more accessible in generating and analyzing massive amounts of data (Figure 14.4). Individual laboratories can now quickly enter the big data era by generating their digital data in-house or through third parties. Unlike the era when specimens were collected and stored in museums or special collections, data storage and analysis now only require access to the cloud.
3. *The expansive breadth of the "omics" framework*. While much focus has been placed on genome technologies, recent advances in assaying the molecules that DNA encodes including transcripts, proteins, metabolites, short and long RNAs, have each gone through significant innovations. Now, omics provides an exhaustive and comprehensive survey of the genome and its array of encoded products. Both non-coding and coding DNA can be analyzed together alongside co-inhabiting microbes and their associated genomes (i.e. microbiomes), transcriptomes, proteomes, and metabolomes (Figure 14.5). Thus, all aspects of a molecular ecosystem can be studied *in digital*. A systems approach provides a multidimensional and increasingly dynamic view of the "static genome" with new innovative technologies being continuously propagated to drive discovery.
4. *Mapping the complexity of phenotypes to DNA*. The field of genetics has historically been based on identifying the genes found in a "mutant" phenotype, i.e. mapping phenotype to genotype in what has been termed as "forward genetics." Systems biology provides a powerful approach that reverses this mapping paradigm from genotype onto phenotype. Units of heredity can be quickly mapped across levels of biological organization in an exhaustive and comprehensive manner through omics. Variations in sequences are now correlated to variable levels of genome-wide transcription, methylation, histone acetylation, and protein and metabolome expression. Thus, a digital DNA code defines the totality of the underlying molecules involved in complex phenotypes.
5. *Using a "systems" approach to study the interactive nature of genetic components*. The molecular biology era allowed biologists to reduce the complexity of phenotypes (e.g.



Note: Genes located on a genome could for a variety of molecules include transcripts, proteins, and metabolites that are part of a complex molecular ecosystem that underlies phenotypes.

Figure 14.5 Omics and the underlying architecture of complex phenotypes

disease) to a single gene, and sometimes a single base pair. However, genes do not work in isolation and, unfortunately, reductionist approaches did not easily lend themselves to the study of epistatic, background, and gene–environmental interactions. Whole genome-encoded omics allowed for new systems-oriented approaches (Hood, Heath, Phelps, & Lin, 2004) giving biologists the unprecedented opportunity to address and identify complex interactions—from pairwise to n-order—between genes, proteins, and other molecules in digital form. Again, it is the ability to quickly and cheaply generate genome-wide sequence data that drives this transformation in our understanding of molecular interactions.

6. *Epigenomics and associating environmental factors to DNA modification*. New discoveries using omics technologies appear to reject the stringent central dogma concerning the irreversibility of information flow (i.e. DNA → mRNA → protein; Crick, 1970). Indeed, the environment appears to play an important role in modifying DNA, gene expression, and the overall phenotype through epigenetics. Since epigenetic modifications are generally enacted genome-wide, omics approaches provide an ideal platform to study environmentally-induced gene regulation across the genome. In particular, biosocial frameworks that integrate genes and the environmental context that individuals are placed in are beginning to view epigenomics as an important mechanism that can explain many aspects of the human condition including disease (Kinsey et al., 2018; Wiese, Rodriguez Escobar, Hsu, Kulathinal, & Hayes-Conroy, 2018).
7. *Embedding of evolutionary history in the code*. Genomics provides biologists with complete records to map sequences back in time in order to study the rich evolutionary history of any species. Using such historical data, we can model inferred changes as well

as document evolutionary patterns that can eventually be used to predict outcomes. This evolutionary lens provides a summary of millions, if not billions, of years of success and failure in debugging an ever-evolving code *in natura*.

6. TOWARDS A FULLY DIGITIZED PLATFORM IN THE BIOLOGICAL SCIENCES

These last two decades have witnessed the digital transformation of biology through unprecedented digital data generation precipitated by the omics revolution. Genomes and their molecular ecosystems can now be easily identified by their long, linear array of A's, T's, G's, and C's, thus, providing an ideal digital platform to map both the complexity and diversity of life. Never has life been recognized as so vast yet so connected.

On one axis, phenotypic complexity, as defined by the rich tapestry of character traits and their underlying molecules, is encoded by specific sequences of nucleotides located on different parts of the genome, i.e. loci. While a subset of these loci code for proteins, others are non-coding and regulatory in nature. Some loci can be environmentally controlled through epigenetic modifications. Many loci interact via their gene products forming complex regulatory networks (Davidson & Levin, 2005). The mapping of these loci to specific traits, i.e. the genotype phenotype map, remains a principal objective in the biological sciences and the omics era has quickly advanced its progress via a digital platform (Figure 14.5).

On the other axis, this same digital platform can also map the diversity of species via their orthologous sequences. Since all known life on our planet share a common 3+ billion old ancestor, sequences can be used to date the origins of different species and the molecules that define them. Further, by applying a comparative approach, we can infer patterns and processes that occurred in the past to garner novel insight about the nature of the code present today. For example, sites that are conserved are implied to be functionally important with deleterious consequences when mutated (i.e. disease). These two complementary axes of life—complexity and diversity—can be digitally connected *in silico*, providing a powerful platform for storage, observation, and analysis of all of biology.

Thus, biology's empirical tradition continues today, but within a digital form. Omics will continue to generate an unprecedented amount of mappable digital data at an exponential rate (Figure 14.4). However, even beyond sequence data, other aspects of our biology are becoming increasingly digitized with opportunities to enable even more connections. New tools and increasing storage capacity are transforming how biologists are utilizing imaged data: cells, tumors, neurons, brain anatomy provide granular phenotypic data that can be stored and analyzed as digitized pixels. Similarly, dynamic behavioral data are being transformed into digital bits and completely analyzed *in silico*. Biology is well on its way to becoming completely digital with important consequences including data democratization and the production of higher rates of innovation.

7. THE DIGITIZATION AND MATHEMATIZATION OF BIOLOGY

One of biology's main goals is to map its DNA code—whether protein-coding, non-coding, regulatory, or an epigenetic mark—onto the detailed phenotypes that make life complex and diverse. As described above, phenotypes are increasingly becoming digitized from omics to pixelated images to time-series binary behavioral data, thus bringing us one step closer to a fully integrated empirical digital platform. There's less of a need for biologists to go to museums, libraries, or even laboratories to investigate their interests as a comprehensive biological platform can be accessed digitally from almost anywhere in the world. Much of this has to do with the proliferation of tens of thousands of open-access databases and databanks (e.g. NCBI, EMBL, DDBJ) that are often expertly curated.

Yet, the true power of this data platform lies in the potential to apply computationally intensive analytics. Today, gene regulatory networks are modeled as digital genetic circuits while their encoded proteins are digitally modeled to help identify drug targets. Our medical history is being collated into digital medical health records and integrated with genomic information to provide collective insight on disease (e.g. Gligorijevic et al., 2016). Fitbit physiological data are being sent to our primary caregivers and automatically surveilled for outliers, often in real-time. Our DNA is being compared to the rest of humanity to identify our genetic ancestry while these same service providers, as part of their business model, use the collective data to map phenotypic traits with unprecedented power to sell to pharmaceutical companies. We can identify mutations in specific cells as well as effortlessly survey their genome-wide effects on various omic landscapes of gene expression. With its massive and ever-growing datasets and capacity to analyze big data, this is certainly the century of digital biology.

Of course, the intricacy of life is the culmination of billions of years of evolutionary tweaking. Due to the connectivity among all life via its shared ancestry, we can unmask the vast history of evolutionary innovations that has made life both possible and impossible. Thus, adding the diversity axis incorporates an exceptionally deep historical data dimension, making our digital platform extremely powerful. Connecting life's rich history fuels predictive biology that links DNA to form, function, and disease. By combining evolutionary theory with big data omics, digital biology provides a ripe platform for mathematization. Whether through population genetics or molecular evolution (e.g. Kimura's neutral theory), or through genetic or biochemical analyses (e.g. Kascer and Burns' metabolic theory), biology has had a long history of applying math to formalize its theory. Now, with unprecedented amounts of digital sequence and phenotypic empirical data, biologists are applying mathematical approaches in their analytics. Life, in all its diversity and complexities, may in fact be calculable. For example, after replicating basic life solely from digital records (Gibson et al., 2010), several laboratories are seeking to produce organisms with minimal genomic content (e.g. Hutchison et al., 2016). Thus, reducing life to its lowest common digital denominator is already underway!

8. THE DIGITAL INNOVATION REVOLUTION IN BIOLOGY: FROM PATTERN TO PROCESS TO PREDICTION

Biologists' understanding of phenotypes has dramatically increased with digitization. The integration of a common molecular vocabulary with other digital technology from imaging

to medical health records and the leveraging of historical insight via evolutionary analysis is transforming biology from a science that was primarily pattern-based, then process-based, to a highly innovative science of prediction.

In describing digital technology innovations, Yoo and colleagues (Yoo, Boland, Lyytinen, & Majchrzak, 2012) prescribe three key traits: (1) the availability of a stable digital technology platform, (2) the emergence of distributed innovations, and (3) the prevalence of combinatorial innovation. The digital biology framework that uses the nucleotides of DNA to connect the molecular ecosystem underlying phenotypes with the entire history of life provides all biologists with a common platform to combine data, analyses, and approaches. Ironically, life itself is the product of continual innovations that involved billions of years of sequence evolution, was widely distributed over phylogenetic and biogeographical space, and implemented recombinational mechanisms to design new features and perfect old ones.

Perhaps the most monumental consequence of biology's digitization is the democratization of *in digital* data. By distributing such data openly and freely and allowing more hands and minds to analyze these data using a growing pool of approaches, the opportunity for innovation is maximized. Of course, one important component of a data analytics infrastructure is the development and maintenance of freely accessible analytical software (e.g. MEGA, Kumar, Stecher, Li, Knyaz, & Tamura, 2018) as well as the integration of new analytical approaches from other fields (e.g. Gligorijevic et al., 2016). For example, all disciplines in biology are quickly heading towards predictive analytics approaches using deep learning techniques. It is certainly an exciting time to be a biologist and the innovation potential generated by digitality drives this excitement.

With the democratization of digital biological data, we see three key opportunities for future work in digital innovation. First, the explosion of different types of biological data from the level of DNA all the way up to the population level and beyond helps advance a new "computational synthesis" in biological research. Each layer of biological data (DNA, proteins, cells, organs, organisms, population, ecosystem, etc.) has its own unique data structure, analytical tools, and models. What is lacking is a comprehensive and integrated data architecture that allows scholars to look at the biological world from a computational system biology level. Just like how generative innovation derived from the layered architecture of the Internet (Benkler, 2006; Yoo et al., 2012), the establishment of such a layered architecture of biological data would certainly accelerate the next generation of innovations in biological and medical research.

Second, as touched upon before, the explosive growth of the digital trace is not limited to genomic data. We currently are witnessing an explosive growth in digital data in other domains. The development of wearable and IOT (Internet of Things) devices offers the opportunity to constantly monitor human behaviors. The explosive growth of social media has produced a trove of digital trace data that capture human behaviors, emotions, interests, and relations. Financial institutions and emerging and e-commerce platforms such as Amazon have collected consumer purchasing transaction data that reflect the behaviors and interests of each individual. With such data, scholars can begin to develop comprehensive multi-modal human databases that entail biological, social, and behavioral data. The development of such multi-modal human data, together with increasingly powerful data analytics tools, allows scholars to pursue radically novel multi-disciplinary inquiry to better understand the human condition and the precise phenotypes that define it. For example, scholars have recently discovered how social networks can influence the diffusion of obesity (Christakis & Fowler,

2007). We can expect more of such novel discovery by connecting how social, behavioral, and technological factors influence the changes in the biological conditions both at the individual and population levels.

Lastly, the emergence of big data with longitudinal digital trace data enables scholars in all fields to explore the nature of "things" from their own field of inquiry which is typically based on a stable and fixed ontology (Delanda, 2002). With the development of digital trace data, we are able to reconceptualize the nature of being not as fixed stable entities, but as a geometry of possibilities that undergoes constant change. Such evolutionary ontological approaches are beginning to gain traction in modern biological research (e.g. coalescent theory, developmental genomics, systems biology). Just as DNA helps define the space of possibilities of any living organism, scholars in other fields can seek to reconceptualize various constructs such as institution, organizations, products, routines, etc., through an evolutionary ontology by defining them through the probabilistic expression of the space of possibilities. The evolutionary ontology, evolutionary theories, and evolutionary methodologies that have been developed in the biological sciences can provide useful templates for other fields interested in adopting a similar evolutionary stance.

9. CONCLUSIONS

Biology is now fully digital with important consequences on the generativity of innovation. Recent paradigm shifts in theory and technology have transformed biology from the descriptive to the analytic and, now, to the predictive. The evolution revolution digitally connects all of life's diversity through common ancestry. The "omics" revolution digitally connects life's complexity across all levels of hierarchical biological organization from genotype to phenotype. This digital transformation is currently embracing technical innovations from the "omics" revolution that treat biology as a series of integrated and interconnected biological network hierarchies from genes to cells to ecosystems, allowing us to map the biological processes that underlie all of earth's biodiversity. The opportunities offered by this emerging digital biological platform include the future development of layered architecture to help map genotypes to phenotypes and the extension of this layered architecture to all fields of inquiry through the combination of biological and social science data and the development of evolutionary ontologies. Treating biology's digital platform as both the object and subject of innovation provides unprecedented power to emerging approaches in phylomedicine and the biosocial and their ability to predict the human condition.

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bentley, D. R., Balasubramanian, S., Swerlow, H. P., Smith, G. P., Milton, J., Brown, C. G., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*, 53–59.

Boland, R., Lyytinen, K., & Yoo, Y. (2007). Wakes of innovation in project networks: The case of digital 3-D representations in architecture, engineering, and construction. *Organization Science*, *18*(4).
Cartwright, J. H.E., Giannerini, S., & González, D. L. (2016). DNA as information: At the crossroads between biology, mathematics, physics and chemistry. *Philosophical Transactions of the Royal Society A*, *374*(2063), 20150071.
Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, *357*, 370–379.
Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*, 561–563.
Darwin, C. (1859). *On the origin of species by means of natural selection*. London: Murray.
Davidson, E., & Levin, M. (2005). Gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 4935.
Delanda, M. (2002). *Intensive science and virtual philosophy*. New York: Bloomsbury.
Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Columbia University Press.
Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *35*(3), 125–129.
Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 14863–14868.
Frankelius, P. (2009). Questioning two myths in innovation literature. *The Journal of High Technology Management Research*, *20*, 40–51.
Gibson, D. G. Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., … Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science, *329*, 52–56.
Gligorijevic, D., Stojanovic, J., Djuric, N., Radosavljevic, V., Grbovic, M., Kulathinal, R. J., & Obradovic, Z. (2016). Large-scale discovery of disease–disease and disease–gene associations. *Scientific Reports*, *6*, 32404.
Hood, L., Heath, J. R., Phelps, M. E., & Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science*, *306*, 640–643.
Hutchison, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280).
Huxley, J. (1942). *Evolution: The modern synthesis*. New York: Harper.
Kinsey, D., Paul, C. P., Taylor, D., Caricchio, R., Kulathinal, R. J., & Hayes-Conroy, A. (2018). The whole lupus: Articulating biosocial interplay in systemic lupus erythematosus epidemiology and population disparities. *Health Place*, *51*, 182–188.
Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549.
Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, time-trees, and divergence times. *Molecular Biology and Evolution*, *34*, 1812–1819.
Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23, 127–128.
Linnaeus, C. (1735). *Systema naturae, sive regna tria naturae systematice proposita per classes, ordines, genera, & species*. Leiden: Haak.
Margulies, M., Egholm M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., … Rothberg, J. M. (2005). Genome sequencing in open microfabricated high density picoliter reactors. *Nature*, *437*, 376–380.
Mayr, E. (1942). *Systematics and the origin of species*. New York: Columbia University Press.
Negroponte, N. (1995). *Being digital*. New York: Vintage Books.
Nirenberg, M. W. & Matthaei, H. J. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, *47*, 1588–1602.
Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A.M., … Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, *309*, 1728–1732.
Simpson, G. G. (1944). *Tempo and mode in evolution*. New York: Columbia University Press.
Stebbins, G. L. (1950). *Variation and evolution in plants*. New York: Columbia University Press.

The International HapMap Consortium (2003). The International HapMap Project. *Nature*, *426*, 789–796.

Um, S., Yoo, Y., Wattal, S., Kulathinal, R., & Zhang, B. (2013). The architecture of generativity in a digital ecosystem: A network biology perspective. Paper presented at the International Conference on Information Systems (ICIS 2013), Reshaping Society through Information Systems Design, Milan, Italy.

Wetterstrand, K. A. (2019). DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). https://www.genome.gov/sequencingcostsdata.

Wiese, D., Rodriguez Escobar, J., Hsu, Y., Kulathinal, R. J., & Hayes-Conroy, A. (2018). The fluidity of biosocial identity and the effects of place, space, and time. Social Science & Medicine, *198*, 46–52.

Yoo, Y., Boland, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, *23*, 1398–1408.