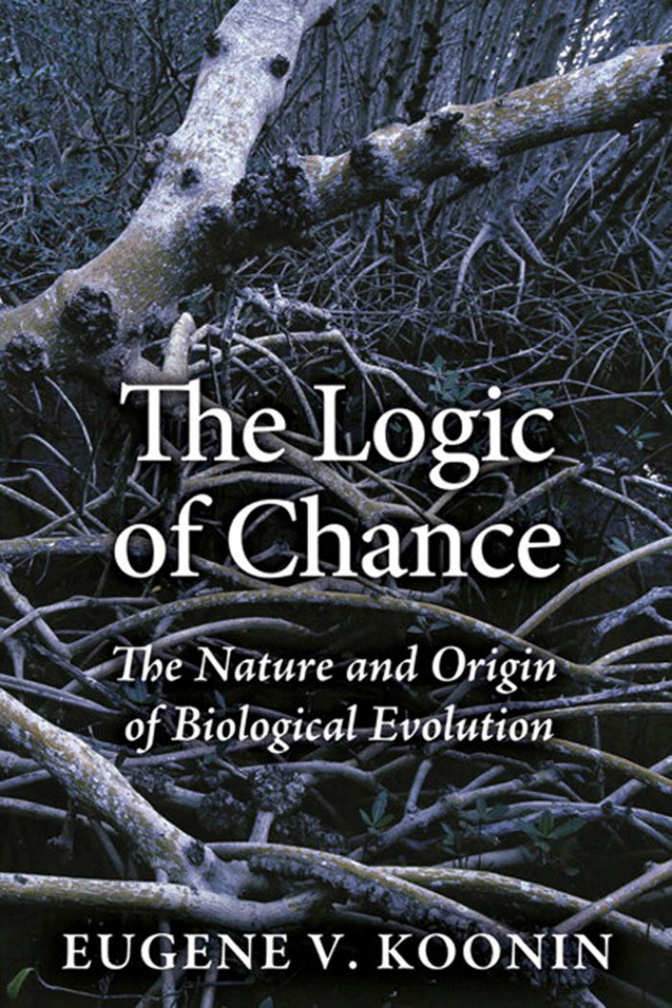

The Logic of Chance

The Nature and Origin of Biological Evolution

EUGENE V. KOONIN

The Logic of Chance:

The Nature and Origin of Biological Evolution

Eugene V. Koonin

Vice President, Publisher: Tim Moore
Associate Publisher and Director of Marketing: Amy Neidlinger
Acquisitions Editor: Kirk Jensen
Editorial Assistant: Pamela Boland
Senior Marketing Manager: Julie Phifer
Assistant Marketing Manager: Megan Graue
Cover Designer: Alan Clements
Managing Editor: Kristy Hart
Project Editor: Betsy Harris
Copy Editor: Krista Hansing Editorial Services, Inc.
Proofreader: Kathy Ruiz
Indexer: Erika Millen
Senior Compositor: Gloria Schurick
Manufacturing Buyer: Dan Uhrig

Publishing as FT Press Science
Upper Saddle River, New Jersey 07458

FT Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact U.S. Corporate and Government Sales, 1-800-382-3419, corpsales@pearsontechgroup.com. For sales outside the U.S., please contact International Sales at international@pearson.com.

Printed in the United States of America

First Printing September 2011

Pearson Education LTD.
Pearson Education Australia PTY, Limited.
Pearson Education Singapore, Pte. Ltd.
Pearson Education Asia, Ltd.
Pearson Education Canada, Ltd.
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan
Pearson Education Malaysia, Pte. Ltd.

ISBN-10: 0-13-254249-8
ISBN-13: 978-0-13-254249-4

Library of Congress Cataloging-in-Publication Data
Koonin, Eugene V.
The logic of chance : the nature and origin of biological evolution / Eugene V. Koonin.—1st ed.
p. cm.
Includes bibliographical references.
ISBN 978-0-13-254249-4 (hardback : alk. paper)
1. Evolutionary genetics. 2. Genomes. 3. Evolution (Biology) I. Title.
QH390.K66 2012
576.8—dc23

2011013400

To my parents

Contents

Preface: Toward a postmodern synthesis of evolutionary biology

The title of this work alludes to four great books: Paul Auster's novel *The Music of Chance* (Auster, 1991); Jacques Monod's famous treatise on molecular biology, evolution, and philosophy, C*hance and Necessity (Le hazard et la necessite)* (Monod, 1972); the complementary book by Francois Jacob, *The Logic of Life* (Jacob, 1993); and, of course, Charles Darwin's *The Origin of Species* (Darwin, 1859). Each of these books, in its own way, addresses the same overarching subject: the interplay of randomness (chance) and regularity (necessity) in life and its evolution.

Only after this book was completed, at the final stage of editing, did I become aware of the fact that the phrase *Logic of Chance* has already been used in a book title by John Venn, an eminent Cambridge logician and philosopher who in 1866 published *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability.* This work is considered to have laid the foundation of the frequency interpretation of probability, which remains the cornerstone of probability theory and statistics to this day (Venn, 1866). He is obviously famous for the invention of the ubiquitous Venn diagrams. I am somewhat embarrassed that I was unaware of John Venn's work when starting this book. On the other hand, I can hardly think of a more worthy predecessor.

My major incentive in writing this book is my belief that, 150 years after Darwin and 40 years after Monod, we now have at hand the data and the concepts to develop a deeper, more complex, and perhaps, more satisfactory understanding of this crucial relationship. I make the case that variously constrained randomness is at the very heart of the entire history of life.

The inspiration for this book has been manifold. The most straightforward incentive to write about the emerging new vision of evolution is the genomic revolution that started in the last decade of the twentieth century and continues to unfold. The opportunity to compare the complete genome sequences of thousands of organisms from all walks of life has qualitatively changed the landscape of evolutionary biology. Our inferences about extinct, ancestral life forms are not anymore the wild guesses they used to be (at least, for organisms with no fossil record). On the contrary, comparing genomes reveals numerous genes that are conserved in major groups of living beings (in some cases, even in all or most of them) and thus gives us a previously unimaginable wealth of information and confidence about the ancestral forms. For example, it is not much of an exaggeration to claim that we have an excellent idea of the core genetic makeup of the last common ancestor of all bacteria that probably lived more than 3.5 billion years ago. The more ancient ancestors are much murkier, but even for those, some features seem to be decipherable. The genomic revolution did more than simply allow credible reconstruction of the gene sets of ancestral life forms. Much more dramatically, it effectively overturned the central metaphor of evolutionary biology (and, arguably, of all biology), the Tree of Life (TOL), by showing that evolutionary trajectories of individual genes are irreconcilably different. Whether the TOL can or should be salvaged—and, if so, in what form—remains a matter of intense debate that is one of the important themes of this book.

Uprooting the TOL is part of what I consider to be a "meta-revolution," a major change in the entire conceptual framework of biology. At the distinct risk of earning the ire of many for associating with a much-maligned cultural thread, I call this major change the transition to a postmodern view of life. Essentially, this signifies the plurality of pattern and process in evolution; the central role of contingency in the evolution of life forms ("evolution as tinkering"); and, more specifically, the demise of (pan)adaptationism as the paradigm of evolutionary biology. Our unfaltering admiration for Darwin notwithstanding, we must relegate the Victorian worldview (including its refurbished versions that flourished in the twentieth century) to the venerable museum halls where it belongs, and explore the consequences of the paradigm shift.

However, this overhaul of evolutionary biology has a crucial counterpoint. Comparative genomics and evolutionary systems biology (such as organism-wide comparative study of gene expression, protein abundance, and other molecular characteristics of the phenotype) have revealed several universal patterns that are conserved across the entire span of cellular life forms, from bacteria to mammals. The existence of such universal patterns suggests that relatively simple theoretical models akin to those employed in statistical physics might be able to explain important aspects of biological evolution; some models of this kind with considerable explanatory power already exist. The notorious "physics envy" that seems to afflict many biologists (myself included) might be soothed by recent and forthcoming theoretical developments. The complementary relationship between the universal trends and the contingency of the specific results of evolution appears central to biological evolution—and the current revolution in evolutionary biology—and this is another central theme of this book.

Another entry point into the sketch of a new evolutionary synthesis that I am trying to develop here is more specific and, in some ways, more personal. I earned my undergraduate and graduate degrees from Moscow State University (in what was then the USSR), in the field of molecular virology. My PhD project involved an experimental study of the replication of poliovirus and related viruses that have a tiny RNA molecule for their genome. I have never been particularly good with my hands, and the time and place were not the best for experimentation because even simple reagents and equipment were hard to obtain. So right after I completed my PhD project, a colleague, Alex Gorbalenya, and I started to veer into an alternative direction of research that, at the time, looked to many like no science at all. It was "sequence gazing"—that is, attempting to decipher the functions of proteins encoded in the genomes of small viruses (the only complete genomes available at the time) from the sequences of their building blocks, amino acids. Nowadays, anyone can rapidly perform such an analysis by using sleek software tools that are freely available on the Internet; naturally, meaningful interpretation of the results still requires thought and skill (that much does not change). Back in 1985, however, there were practically no computers and no software. Nevertheless, with our computer science colleagues, we managed to develop some rather handy programs (encoded at the time on punch cards). Much of the analysis was done by hand (and eye). Against all odds, and despite some missed opportunities and a few unfortunate errors, our efforts over the next five years were remarkably fruitful. Indeed, we managed to transform the functional maps of those small genomes from mostly unchartered territory to fairly rich "genomescapes" of functional domains. Most of these predictions have been subsequently validated by experiment, and some are still in the works (bench experimentation is much slower than computational analysis). I believe that our success was mostly due to the early realization of the strikingly simple but surprisingly powerful basic principle of evolutionary biology: When a distinct sequence motif is conserved over a long evolutionary span, it must be functionally important, and the higher the degree of conservation, the more important the function. This common-sense principle that is of course rooted in the theory of molecular evolution has served our purposes exceedingly well and, I believe, converted me into an evolutionary biologist for the rest of my days. What I mean is not so much theoretical knowledge, but rather an indelible feeling of the absolute centrality and essentiality of evolution in biology. I am inclined to reword the famous dictum of the great evolutionary geneticist Theodosius Dobzhansky ("Nothing in biology makes sense except in the light of evolution") (Dobzhansky, 1973) in an even more straightforward manner: *Biology is evolution*.

In those early days of evolutionary genomics, Alex and I often talked about the possibility that our beloved small RNA viruses could be direct descendants of some of the earliest life forms. After all, they were tiny and simple genetic systems, with only one type of nucleic acid involved, and their replication was directly linked to expression through the translation of the genomic RNA. Of course, this was late-night talk with no direct relevance to our daytime effort on mapping the functional domains of viral proteins. However, I believe that, 25 years and hundreds of diverse viral and host genomes later, the idea that viruses (or virus-like genetic elements) might have been central to the earliest stages of life's evolution has grown from a fanciful speculation to a concept that is compatible with a wealth of empirical data. In my opinion, this is the most promising line of thought and analysis in the study of the earliest stages of the evolution of life.

So these are the diverse conceptual threads that, to me, unexpectedly converge on the growing realization that our understanding of evolution—and, with it, the very nature of biology—have forever departed from the prevailing views of the twentieth century that today look both rather naïve and somewhat dogmatic. At some point, the temptation to try my hand in tying together these different threads into a semblance of a coherent picture became irresistible, hence this book.

Some of the inspiration came from outside of biology, from the recent astounding and enormously fascinating developments in physical cosmology. These developments not only put cosmology research squarely within the physical sciences, but completely overturn our ideas about the way the world is, particularly, the nature of randomness and necessity. When it comes to the boundaries of biology, as in the origin of life problem, this new worldview cannot be ignored. Increasingly, physicists and cosmologists pose the question "Why is there something in the world rather than nothing?" not as a philosophical problem, but as a physical problem, and explore possible answers in the form of concrete physical models. It is hard not to ask the same about the biological world, yet at more than one level: Why is there life at all rather than just solutions of ions and small molecules? And, closer to home, even assuming that there is life, why are there palms and butterflies, and cats and bats, instead of just bacteria? I believe that these questions can be given a straightforward, scientific slant, and plausible, even if tentative, answers seem to be emerging.

Recent advances in high-energy physics and cosmology inspired this book in more than only the direct scientific sense. Many of the leading theoretical physicists and cosmologists have turned out to be gifted writers of popular and semipopular books (one starts to wonder whether there is some intrinsic link between abstract thinking at the highest level and literary talent) that convey the excitement of their revelations about the universe with admirable clarity, elegance, and panache. The modern wave of such literature that coincides with the revolution in cosmology started with Stephen Hawking's 1988 classic *A Brief History of Time* (Hawking, 1988). Since then, dozens of fine diverse books have appeared. The one that did the most to transform my own view of the world is the wonderful and short *Many Worlds in One,* by Alex Vilenkin (Vilenkin, 2007), but equally excellent treatises by Steven Weinberg (Weinberg, 1994), Alan Guth (Guth, 1998a), Leonard Susskind (Susskind, 2006b), Sean Carroll (Carroll, 2010), and Lee Smolin (in a controversial book on "cosmic natural selection"; Smolin, 1999) were of major importance as well. These books are far more than brilliant popularizations: Each one strives to present a coherent, general vision of both the fundamental nature of the world and the state of the science that explores it. Each of these visions is unique, but in many aspects, they are congruent and complementary. Each is deeply rooted in hard science but also contains elements of extrapolation and speculation, sweeping generalizations, and, certainly, controversy. The more I read these books and pondered the implications of the emerging new worldview, the more strongly was I tempted to try something like that in my own field of evolutionary biology. At one point, while reading Vilenkin's book, it dawned on me that there might be a direct and crucial connection between the new perspective on probability and chance imposed by modern cosmology and the origin of life—or, more precisely, the origin of biological evolution. The overwhelming importance of chance in the emergence of life on Earth suggested by this line of enquiry is definitely unorthodox and is certain to make many uncomfortable, but I strongly felt that it could not be disregarded if I wanted to be serious about the origin of life.

This book certainly is a personal take on the current state of evolutionary biology as viewed from the vantage point of comparative genomics and evolutionary systems biology. As such, it necessarily blends established facts and strongly supported theoretical models with conjecture and speculation. Throughout the book, I try to distinguish between the two as best I can. I intended to write the book in the style of the aforementioned excellent popular books in physics, but the story took a life of its own and refused to be written that way. The result is a far more scientific, specialized text than originally intended, although still a largely nontechnical one, with only a few methods described in an oversimplified manner. An important disclaimer: Although the book addresses diverse aspects of evolution, it remains a collection of chapters on selected subjects and is by no account a comprehensive treatise. Many important and popular subjects, such as the origin of multicellular organisms or evolution of animal development, are completely and purposefully ignored. As best I could, I tried to stick with the leitmotif of the book, the interplay between chance and nonrandom processes. Another thorny issue has to do with citations: An attempt to be, if not comprehensive, then at least reasonably complete, would require thousands of references. I gave up on any such attempt from the start, so the reference list at the end is but a small subset of the relevant citations, and the selection is partly subjective. My sincere apologies to all colleagues whose important work is not cited.

All these caveats and disclaimers notwithstanding, it is my hope that the generalizations and ideas presented here will be of interest to many fellow scientists and students—not only biologists, but also physicists, chemists, geologists, and others interested in the evolution and origin of life.

1

The fundamentals of evolution: Darwin and Modern Synthesis

In this chapter and the next, I set out to provide a brief summary of the state of evolutionary biology before the advent of comparative genomics in 1995. Clearly, the task of distilling a century and a half of evolutionary thought and research into two brief, nearly nontechnical chapters is daunting, to put it mildly. Nevertheless, I believe that we can start by asking ourselves a straightforward question: What is the take-home message from all those decades of scholarship? We can garner a concise and sensible synopsis of the pregenomic evolutionary synthesis even while inevitably omitting most of the specifics.

I have attempted to combine history and logic in these first two chapters, but some degree of arbitrariness is unavoidable. In this chapter, I trace the conceptual development of evolutionary biology from Charles Darwin's *On the Origin of Species* to the consolidation of Modern Synthesis in the 1950s. Chapter 2 deals with the concepts and discoveries that affected the understanding of evolution between the completion of Modern Synthesis and the genomic revolution of the 1990s.

Darwin and the first evolutionary synthesis: Its grandeur, constraints, and difficulties

It is rather strange to contemplate the fact that we have just celebrated the 150th anniversary of the first publication of Darwin's *On the Origin of Species* (Darwin, 1859) and the 200th jubilee of Darwin himself. Considering the profound and indelible effect that *Origin* had on all of science, philosophy, and human thinking in general (far beyond the confines of biology), 150 years feels like a very short time.

What was so dramatic and important about the change in our worldview that Darwin prompted? Darwin did not discover evolution (as sometimes claimed overtly but much more often implied, especially in popular accounts and public debates). Many scholars before him, including luminaries of their day, believed that organisms changed over time in a nonrandom manner. Even apart from the great (somewhat legendary) Greek philosophers Empedokles, Parmenides, and Heraclites, and their Indian contemporaries who discussed eerily prescient ideas (even if, oddly for us, combined with mythology) on the processes of change in nature, Darwin had many predecessors in the eighteenth and early nineteenth centuries. In later editions of *Origin,* Darwin acknowledged their contributions with characteristic candor and generosity. Darwin's own grandfather, Erasmus, and the famous French botanist and zoologist Jean-Bapteste Lamarck (Lamarck, 1809) discussed evolution in lengthy tomes.[1] Lamarck even had a coherent concept of the mechanisms that, in his view, perpetuated these changes. Moreover, Darwin's famed hero, teacher, and friend, the great geologist Sir Charles Lyell, wrote about the "struggle for existence" in which the more fecund will always win. And, of course, it is well known that Darwin's younger contemporary, Alfred Russel Wallace, simultaneously proposed essentially the same concept of evolution and its mechanisms.

However, the achievements of all these early evolutionists notwithstanding, it was Darwin who laid the foundation of modern biology and forever changed the scientific outlook of the world in *Origin.* What made Darwin's work unique and decisive? Looking back at his feat from our 150-year distance, three breakthrough generalizations seem to stand out:

1. Darwin presented his vision of evolution within a completely naturalist and rationalist framework, without invoking any teleological forces or drives for perfection (or an outright creator) that theorists of his day commonly considered.
2. Darwin proposed a specific, straightforward, and readily understandable mechanism of evolution that is interplay

between heritable variation and natural selection, collectively described as the survival of the fittest.

3. Darwin boldly extended the notion of evolution to the entire history of life, which he believed could be adequately represented as a grand tree (the famous single illustration of *Origin*), and even postulated that all existing life forms shared a single common ancestor.

Darwin's general, powerful concept stood in stark contrast to the evolutionary ideas of his predecessors, particularly Lamarck and Lyell, who contemplated mostly, if not exclusively, evolutionary change within species. Darwin's fourth great achievement was not purely scientific, but rather presentational. Largely because of a well-justified feeling of urgency caused by competition with Wallace, Darwin presented his concept in a brief and readable (even for prepared lay readers), although meticulous and carefully argued, volume. Thanks to these breakthroughs, Darwin succeeded in changing the face of science rather than just publishing another book. Immediately after *Origin* was published, most biologists and even the general educated public recognized it as a credible naturalist account of how the observed diversity of life could have come about, and this was a dynamic foundation to build upon.[2]

Considering Darwin's work in a higher plane of abstraction that is central to this book, it is worth emphasizing that Darwin seems to have been the first to establish the crucial interaction between chance and order (necessity) in evolution. Under Darwin's concept, variation is (nearly) completely random, whereas selection introduces order and creates complexity. In this respect, Darwin is diametrically opposed to Lamarck, whose worldview essentially banished chance. We return to this key conflict of worldviews throughout the book.

Indeed, with all due credit given to his geologist and early evolutionary biologist predecessors, Darwin was arguably the first scholar to prominently bring the possibility of evolutionary change (and, by implication, origin) of the entire universe into the realm of natural phenomena that are subject to rational study. Put another way, Darwin initiated the scientific study of the *time arrow*—that is, time-asymmetrical, irreversible processes. By doing so, he prepared the ground not only for all further development of biology, but also for the advent of modern physics. I believe that the great physicist Ludwig Boltzmann, the founder of statistical thermodynamics and the author of the modern concept of entropy, had good reason to call Darwin a "great physicist," paradoxical as this might seem, given that Darwin knew precious little about actual physics and mathematics. Contemporary philosopher Daniel Dennett may have had a point when he suggested that Darwin's idea of natural selection might be the single greatest idea ever proposed (Dennett, 1996).

Certainly, Darwin's concept of evolution at the time *Origin* was published and at least through the rest of the nineteenth century faced severe problems that greatly bothered Darwin and, at times, appeared insurmountable to many scientists. The first substantial difficulty was the low estimate of the age of Earth that prevailed in Darwin's day. Apart from any creation myth, the best estimates by nineteenth-century physicists (in particular, Lord Kelvin) were close to 100 million years, a time span that was deemed insufficient for the evolution of life via the Darwinian route of gradual accumulation of small changes. Clearly, that was a correct judgment—the 100 million years time range is far too short for the modern diversity of life to evolve, although no one in the nineteenth century had a quantitative estimate of the rate of Darwinian evolution. The problem was resolved 20 years after Darwin's death. In the beginning of the twentieth century, when radioactivity was discovered, scientists calculated that cooling of the Earth from its initial hot state would take billions of years, just about the time Darwin thought would be required for the evolution of life by natural selection.

The second, more formidable problem has to do with the mechanisms of heredity and the so-called Jenkin nightmare. Because the concept of discrete hereditary determinants did not exist in Darwin's time (outside the obscure articles of Mendel), it was unclear how an emerging beneficial variation could survive through generations and get fixed in evolving populations without being diluted and perishing. Darwin apparently did not think of this problem at the time he wrote *Origin;* an unusually incisive reader, an engineer named Jenkin, informed Darwin of this challenge to his theory. In retrospect, it is difficult to understand how Darwin (or Jenkin or Huxley) did not think of a Mendelian solution. Instead, Darwin came up with a more extravagant concept of heredity, the so-called pangenesis, which even he himself did not seem to take quite seriously. This problem was resolved by the (re)birth of genetics, although the initial implications for Darwinism[3] were unexpected (see the next section).

The third problem that Darwin fully realized and brilliantly examined was the evolution of complex structures (*organs,* in Darwin's terms) that require assembly of multiple parts to perform their function. Such complex organs posed the classic puzzle of evolutionary biology that, in the twentieth century, has been evocatively branded 'irreducible complexity.'[4] Indeed, it is not immediately clear how selection could enact the evolution of such organs under the assumption that individual parts or partial assemblies are useless. Darwin tackled this problem head-on in one of the most famous passages of *Origin,* the scenario of evolution of the eye. His proposed solution was logically impeccable, plausible, and ingenious: Darwin posited that complex organs do evolve through a series of intermediate stages, each of which performs a partial function related to the ultimate function of the evolving complex organ. Thus, the evolution of the eye, according to Darwin, starts with a simple light-sensing patch and proceeds through primitive eye-like structures of incrementally increasing utility to full-fledged, complex eyes of arthropods and vertebrates. It is worth noting that primitive light-sensing structures resembling those Darwin postulated on general grounds have been subsequently discovered, at least partially validating his scenario and showing that, in this case, the irreducibility of a complex organ is illusory. However, all the brilliance of Darwin's scheme notwithstanding, it should be taken for what it is: a partially supported speculative scenario for the evolution of one particular complex organ. Darwin's account shows one possible trajectory for the evolution of complexity but does not solve this major problem in general. Evolution of complexity at different levels is central to understanding biology, so we revisit it on multiple occasions throughout this book.

The fourth area of difficulty for Darwinism is, perhaps, the deepest. This major problem has to do with the title and purported main subject of Darwin's book, the origin of species and, more generally, large-scale evolutionary events that are now collectively denoted as macroevolution. In a rather striking departure from the title of the book, all indisputable examples of evolution that Darwin presented involve the emergence of new varieties within a species, not new species let alone higher taxa. This difficulty persisted long after

Darwin's death and exists even now, although it was mitigated first by the progress of paleontology, then by developments in the theory of speciation supported by biogeographic data, and then, most convincingly, by comparative genomics (see Chapters 2 and 3). Much to his credit, and unlike detractors of evolution up to this day, Darwin firmly stood his ground in the face of all difficulties, thanks to his unflinching belief that, incomplete as his theory might be, there was no rational alternative. The only sign of Darwin's vulnerability was the inclusion of the implausible pangenesis model in later editions of *Origin,* as a stop-gap measure to stave off the Jenkin nightmare.

Genetics and the "black day" of Darwinism

An urban legend tells that Darwin had read Mendel's paper but found it uninspiring (perhaps partly because of his limited command of German). It is difficult to tell how different the history of biology would have been if Darwin had absorbed Mendel's message, which seems so elementary to us. Yet this was not to be.

Perhaps more surprisingly, Mendel himself, although obviously well familiar with the *Origin,* did not at all put his discovery into a Darwinian context. That vital connection had to await not only the rediscovery of genetics at the brink of the twentieth century, but also the advent of population genetics in the 1920s. The rediscovery of Mendelian inheritance and the birth of genetics should have been a huge boost to Darwinism because, by revealing the discreteness of the determinants of inheritance, these discoveries eliminated the Jenkin nightmare. It is therefore outright paradoxical that the original reaction of most biologists to the discovery of genes was that genetics made Darwin's concept irrelevant, even though no serious scientist would deny the reality of evolution. The main reason genetics was deemed incompatible with Darwinism was that the founders of genetics, particularly Hugo de Vries, the most productive scientist among the three rediscoverers of Mendel laws, viewed mutations of genes as abrupt, saltational hereditary changes that ran counter to Darwinian gradualism. These mutations were considered to be an inalienable feature of Darwinism, in full accord with *Origin.* Accordingly, de Vries viewed his mutational theory of evolution as "anti-Darwinian." So Darwin's centennial jubilee and the 50th anniversary of the *Origin* in 1909 were far from triumphant, even as genetic research surged and Wilhelm Johansson introduced the term *gene* that very year.

Population genetics, Fisher's theorem, fitness landscapes, drift, and draft

The foundations for the critically important synthesis of Darwinism and genetics were set in the late 1920s and early 1930s by the trio of outstanding theoretical geneticists: Ronald Fisher, Sewall Wright, and J. B. S. Haldane. They applied rigorous mathematics and statistics to develop an idealized description of the evolution of biological populations. The great statistician Fisher apparently was the first to see that, far from damning Darwinism, genetics provided a natural, solid foundation for Darwinian evolution. Fisher summarized his conclusions in the seminal 1930 book *The Genetical Theory of Natural Selection* (Fisher, 1930), a tome second perhaps only to Darwin's *Origin* in its importance for evolutionary biology.[5] This was the beginning of a spectacular revival of Darwinism that later became known as *Modern Synthesis* (a term mostly used in the United States) or *neo-Darwinism* (in the British and European traditions).

It is neither necessary nor practically feasible to present here the basics of population genetics.[6] However, several generalizations that are germane to the rest of the discussion of today's evolutionary biology can be presented succinctly. Such a summary, even if superficial, is essential here. Basically, the founders of population genetics realized the plain fact that evolution does not affect isolated organisms or abstract species, but rather affects concrete groups of interbreeding individuals, termed populations. The size and structure of the evolving population largely determines the trajectory and outcome of evolution. In particular, Fisher formulated and proved the fundamental theorem of natural selection (commonly known as Fisher's theorem), which states that the intensity of selection (and, hence, the rate of evolution due to selection) is proportional to the magnitude of the standing genetic variation in an evolving population, which, in turn, is proportional to the effective population size.

Box 1-1 gives the basic definitions and equations that determine the effects of mutation and selection on the elimination or fixation of mutant alleles, depending on the effective population size. The qualitative bottom line is that, given the same mutation rate, in a population with a large effective size, selection is intense. In this case, even mutations with a small positive selection coefficient ("slightly" beneficial mutations) quickly come to fixation. On the other hand, mutations with even a small negative selection coefficient (slightly deleterious mutations) are rapidly eliminated. This effect found its rigorous realization in Fisher's theorem.

Box 1-1: The fundamental relationships defining the roles of selection and drift in the evolution of populations

Nearly neutral evolution dominated by drift

$1/Ne >> |s|$

Evolution dominated by selection

$1/Ne << |s|$

Mixed regime, with both drift and selection important

$1/Ne \approx |s|$

Ne: effective population size (typically, substantially less than the number of individuals in a population because not all individuals produce viable offspring)

s: selection coefficient or fitness effect of mutation:

$s = F_A - F_a$

F_A, F_a: fitness values of two alleles of a gene

$s>0$: beneficial mutation

$s<0$: deleterious mutation

A corollary of Fisher's theorem is that, assuming that natural selection drives all evolution, *the mean fitness of a population cannot decrease during evolution* (if the population is to survive, that is). This is probably best envisaged using the imagery of a *fitness landscape*, which was first introduced by another founding father of population genetics, Sewall Wright. When asked by his mentor to present the results of his mathematical analysis of selection in a form accessible to biologists, Wright came up with this extremely lucky image. The appeal and simplicity of the landscape representation of fitness evolution survive to this day and have stimulated numerous subsequent studies that have yielded much more sophisticated and less intuitive theories and versions of fitness landscapes, including multidimensional ones (Gavrilets, 2004).[7] According to Fisher's theorem, a population that evolves by selection only (technically, a population of an infinite size—infinite populations certainly do not actually exist, but this is convenient abstraction routinely used in population genetics) can never move downhill on the fitness landscape (see Figure 1-1). It is easy to realize that a fitness landscape, like a real one, can have many different shapes. Under certain special circumstances, the landscape might be extremely smooth, with a single peak corresponding to the global fitness maximum (sometimes this is poetically called the Mount Fujiyama landscape; see Figure 1-1A). More realistically, however, the landscape is rugged, with multiple peaks of different heights separated by valleys (see Figure 1-1B). As formally captured in Fisher's theorem (and much in line with Darwin), a population evolving by selection can move only uphill and so can reach only the local peak, even if its height is much less than the height of the global peak (see Figure 1-1B). According to Darwin and Modern Synthesis, movement across valleys is forbidden because it would involve a downhill component. However, the development of population genetics and its implications for the evolutionary process changed this placid picture because of *genetic drift,* a key concept in evolutionary biology that Wright also introduced.

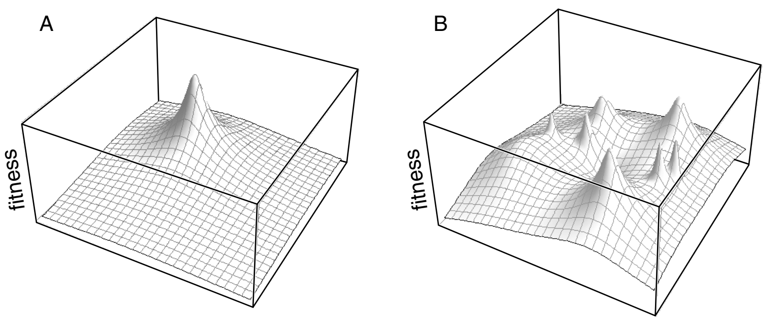

Figure 1-1 Fitness landscapes: the Mount Fujiyama landscape with a single (global) fitness peak and a rugged fitness landscape.

As emphasized earlier, Darwin recognized a crucial role of chance in evolution, but that role was limited to one part of the evolutionary process only: the emergence of changes (mutations, in the modern parlance). The rest of evolution was envisaged as a deterministic domain of necessity, with selection fixing advantageous mutations and the rest of mutations being eliminated without any long-term consequence. However, when population dynamics entered the picture, the situation changed dramatically. The founders of quantitative population genetics encapsulated in simple formulas the dependence of the intensity of selection on population size and mutation rate (see Box 1-1 and Figure 1-2). In a large population with a high mutation rate, selection is effective, and even a slightly advantageous mutation is fixed with near certainty (in an infinite population, a mutation with an infinitesimally small positive selection coefficient is fixed deterministically). Wright realized that a small population, especially one with a low mutation rate, is quite different. Here random genetic drift plays a crucial role in evolution through which neutral or even deleterious (but, of course, nonlethal) mutations are often fixed by sheer chance. Clearly, through drift, an evolving population can violate the principle of upward-only movement in the fitness landscape and might slip down (see Figure 1-2).[8] Most of the time, this results in a downward movement and subsequent extinction, but if the valley separating the local peak from another, perhaps taller one is narrow, then crossing the valley and starting a climb to a new, perhaps taller summit becomes possible (see Figure 1-2). The introduction of the notion of drift into the evolutionary narrative is central to my story. Here chance enters the picture at a new level: Although Darwin and his immediate successors saw the role of chance in the emergence of heritable change (mutations), drift introduces chance into the next phase—namely, the fixation of these changes—and takes away some of the responsibility from selection. I explore just how important the role of drift is in different situations during evolution throughout this book.

John Maynard Smith and, later, John Gillespie developed the theory and computer models to demonstrate the existence of a distinct

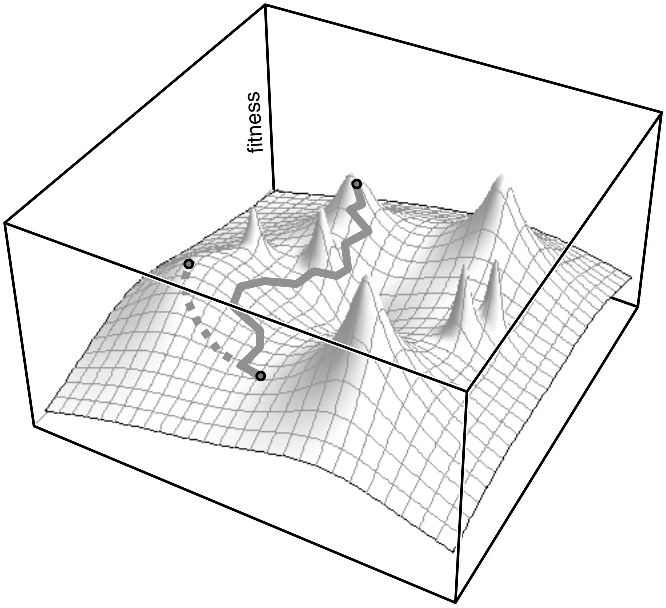

Figure 1-2 Trajectories on a rugged fitness landscape. The dotted line is an evolutionary trajectory at a high effective population size. The solid line is an evolutionary trajectory at a low effective population size.

mode of neutral evolution that is only weakly dependent on the effective population size and that is relevant even in infinite populations with strong selection. This form of neutral fixation of mutations became known as *genetic draft* and refers to situations in which one or more neutral or even moderately deleterious mutations spread in a population and are eventually fixed because of the linkage with a beneficial mutation: The neutral or deleterious alleles spread by *hitchhiking* with the linked advantageous allele (Barton, 2000). Some population-genetic data and models seem to suggest that genetic draft is even more important for the evolution in sexual populations than drift. Clearly, genetic draft is caused by combined effects of natural selection and neutral variation at different genomic sites and, unlike drift, can occur even in effectively infinite populations (Gillespie, 2000).

Genetic draft may allow even large populations to fix slightly deleterious mutations and, hence, provides them with the potential to cross valleys on the fitness landscape.

Positive and purifying (negative) selection: Classifying the forms of selection

Darwin thought of natural selection primarily in terms of fixation of beneficial changes. He realized that evolution weeded out deleterious changes, but he did not interpret this elimination on the same plane with natural selection. In the course of the evolution of Modern Synthesis, the notion of selection was expanded to include “purifying” (negative) selection; in some phases of evolution, this turns out to be more common (orders of magnitude more common, actually) than “Darwinian,” positive selection. Essentially, purifying selection is the default process of elimination of the unfit. Nevertheless, defining this process as a special form of selection seems justified and important because it emphasizes the crucial role of elimination in shaping (constraining) biological diversity at all levels. Simply put, variation is permitted only if it does not confer a significant disadvantage on any surviving variant. To what extent these constraints actually limit the space available for evolution is an interesting and still open issue, and I touch on this later (see in particular Chapters 3, 8, and 9).

A subtle but substantial difference exists between purifying selection and *stabilizing selection,* which is a form of selection defined by its effect on frequency distributions of trait values. These forms include stabilizing selection that is based primarily on purifying selection, directional selection driven by positive (Darwinian) selection, and the somewhat more exotic regimes of disruptive and balancing selection that result from combinations of multiple constraints (see Figure 1-3).

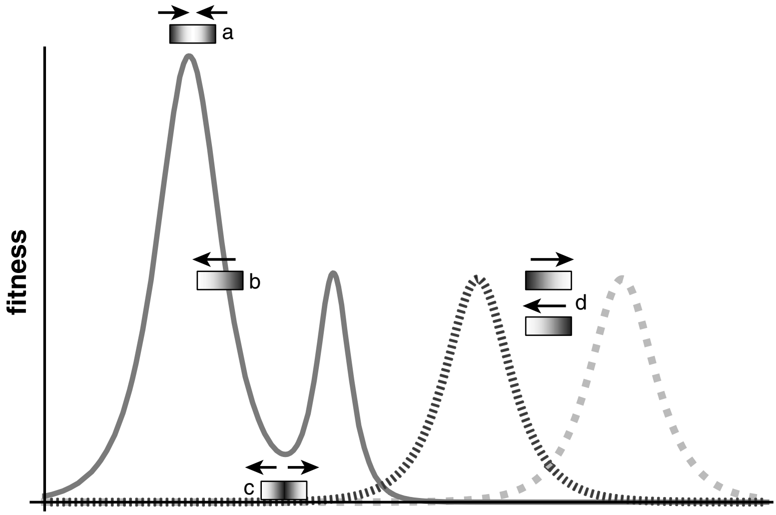

Figure 1-3 Four distinct forms of selection in an evolving population: (A) Stabilizing selection (fitness landscape represented by solid line); (B) Directional selection (fitness landscape represented by solid line); (C) Disruptive selection (fitness landscape represented by solid line); (D) Balancing selection (fitness landscape changes periodically by switching between two dotted lines).

Modern Synthesis

The unification of Darwinian evolution and genetics achieved primarily in the seminal studies of Fisher, Wright, and Haldane prepared the grounds for the Modern Synthesis of evolutionary biology. The phrase itself comes from the eponymous 1942 book by Julian Huxley (Huxley, 2010), but the conceptual framework of Modern Synthesis is considered to have solidified only in 1959, during the centennial celebration of *Origin.* The new synthesis itself was the work of many outstanding scientists. The chief architects of Modern Synthesis were arguably experimental geneticist Theodosius Dobzhansky, zoologist Ernst Mayr, and paleontologist George Gaylord Simpson. Dobzhansky's experimental and field work with the fruit fly *Drosophila melanogaster* provided the vital material support to the theory of population genetics and was the first large-scale experimental validation of the concept of natural selection. Dobzhansky's book *Genetics and the Origin of Species* (Dobzhansky, 1951) is the principal manifesto of Modern Synthesis, in which he narrowly defined evolution as "change in the frequency of an allele within a gene pool." Dobzhansky also famously declared that *nothing in biology makes sense except in the light of evolution*[9] (see more about "making sense" in Appendix A). Ernst Mayr, more than any other scientist, is to be credited with an earnest and extremely influential attempt at a theoretical framework for Darwin's quest, the origin of species. Mayr formulated the so-called biological concept of species, according to which speciation occurs when two (sexual) populations are isolated from each other for a sufficiently long time to ensure irreversible genetic incompatibility (Mayr, 1963). Simpson reconstructed the most comprehensive (in his time) picture of the evolution of life based on the fossil record (Simpson, 1983). Strikingly, Simpson recognized the prevalence of stasis in the evolution of most species and the abrupt replacement of dominant species. He also introduced the concept of quantum evolution, which presaged the punctuated equilibrium concept of Stephen Jay Gould and Niles Eldredge (see Chapter 2).

The consolidation of Modern Synthesis in the 1950s was a somewhat strange process that included remarkable "hardening" (Gould's word) of the principal ideas of Darwin (Gould, 2002). Thus, the doctrine of Modern Synthesis effectively left out Wright's concept of random genetic drift and its evolutionary importance, and remains uncompromisingly pan-adaptationist. Rather similarly, Simpson himself gave up the idea of quantum evolution, so gradualism remained one of the undisputed pillars of Modern Synthesis. This "hardening" shaped Modern Synthesis as a relatively narrow, in some ways dogmatic conceptual framework.

To proceed with the further discussion of the evolution of evolutionary biology and its transformation in the age of genomics, it seems necessary to succinctly recapitulate the fundamental principles of evolution that Darwin first formulated, the first generation of evolutionary biologists then amended, and Modern Synthesis finally codified. We return to each of these crucial points throughout the book.

1. Undirected, random variation is the main process that provides the material for evolution. Darwin was the first to allow chance as a major factor into the history of life, and this was arguably one of his greatest insights. Darwin also allowed a subsidiary role for directed, Lamarckian-type variation, and he tended to give these mechanisms more weight in later editions of *Origin*. Modern

Synthesis, however, is adamant in its insistence on random mutations being the only source of evolutionarily relevant variability.

2. Evolution proceeds by fixation of rare beneficial variations and elimination of deleterious variations: This is the process of natural selection that, along with random variation, is the principal driving force of evolution, according to Darwin and Modern Synthesis. Natural selection, which is obviously akin to and inspired by the "invisible hand" of the market that ruled economy according to Adam Smith, was the first mechanism of evolution ever proposed that was simple and plausible and that did not require any mysterious innate trends. As such, this was Darwin's second key insight. Sewall Wright emphasized that chance could play a substantial role in the fixation of changes during evolution rather than only in their emergence, via genetic drift that entails random fixation of neutral or even moderately deleterious changes. Population-genetic theory indicates that drift is particularly important in small populations that go through bottlenecks. Genetic draft (hitchhiking) is another form of stochastic fixation of nonbeneficial mutations. However, *Modern Synthesis in its "hardened" form effectively rejected the role of stochastic processes in evolution beyond the origin of variation and adhered to a purely adaptationist (pan-adaptationist) view of evolution.* This model inevitably leads to the concept of "progress," gradual improvement of "organs" during evolution. Darwin endorsed this idea as a general trend, despite his clear understanding that organisms are less than perfectly adapted, as strikingly exemplified by rudimentary organs, and despite his abhorrence of any semblance of an innate strive for perfection of the Lamarckian ilk. Modern Synthesis shuns progress as an anthropomorphic concept but nevertheless maintains that evolution, in general, proceeds from simple to complex forms.

3. The beneficial changes that are fixed by natural selection are infinitesimally small (in modern parlance, the evolutionarily relevant mutations are supposed to have infinitesimally small fitness effects), so evolution occurs via the gradual accumulation of these tiny modifications. Darwin insisted on *strict gradualism* as an essential staple of his theory: "Natural selection can act only by the preservation and accumulation of infinitesimally small inherited modifications, each profitable to the preserved being. ...If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down." (*Origin of Species,* Chapter 6). Even some contemporaries of Darwin believed that was an unnecessary stricture on the theory. In particular, the early objections of Thomas Huxley are well known: Even before the publication of *Origin,* Huxley wrote to Darwin, "You have loaded yourself with an unnecessary difficulty in adopting *Natura non facit saltum* so unreservedly" (http://aleph0.clarku.edu/huxley/). Disregarding these early warnings and even Simpson's concept of quantum evolution, Modern Synthesis uncompromisingly embraced gradualism.

4. An aspect of the classic evolutionary biology that is related to but also distinct from the principled gradualism is *uniformitarianism* (absorbed by Darwin from Lyell's geology). This is the belief that the evolutionary processes have remained essentially the same throughout the history of life.

5. This key principle is logically linked to gradualism and uniformitarianism: *Macroevolution* (the origin of species and higher taxa), is governed by the same mechanisms as *microevolution* (evolution within species). Dobzhansky, with his definition of evolution as the change of allele frequencies in populations, was the chief proponent of this principle. Darwin did not use the terms *microevolution* and *macroevolution;* nevertheless, the sufficiency of intraspecies processes to explain the origin of species and, more broadly, the entire evolution of life can be considered the central Darwinian axiom (or perhaps a fundamental theorem, but one for which Darwin did not have even an inkling of the proof). It seems reasonable to speak of this principle as "generalized uniformitarianism": *The processes of evolution are the same not only throughout the history of life, but also at different levels of evolutionary transformation, including major transitions.* The conundrum of microevolution versus macroevolution is, in some ways, the fulcrum of evolutionary biology, so we revisit it repeatedly throughout this book.

6. Evolution of life can be accurately represented by a "great tree," as emphasized by the only illustration in *Origin* (in Chapter 4). Darwin introduced the Tree of Life (TOL) only as a general concept and did not attempt to investigate its actual branching order. The tree was populated with actual life forms, to the best of the knowledge at the time, by the chief German follower of Darwin, Ernst Haeckel. The founders of Modern Synthesis were not particularly interested in the TOL, but they certainly embraced it as a depiction of the evolution of animals and plants that the fossil record amply supported in the twentieth century. By contrast, microbes that were increasingly recognized as major ecological agents remained effectively outside the scope of evolutionary biology.
7. A corollary of the single TOL concept deserves the status of a separate principle: All extant diversity of life forms evolved from a single common ancestor (or very few ancestral forms, under Darwin's cautious formula in Chapter 14 of *Origin*; see Darwin, 1859). Many years later, this has been dubbed the Last Universal Common (Cellular) Ancestor (LUCA). For the architects of Modern Synthesis, the existence of LUCA was hardly in doubt, but they did not seem to consider elucidation of its nature a realistic or important scientific goal.

Synopsis

In his book *On the Origin of Species,* Charles Darwin meticulously collected evidence of temporal change that permeates the world of living beings and proposed for the first time a plausible mechanism of evolution: natural selection. Evolution by natural selection certainly is one of the most consequential concepts ever developed by a scientist and even has been deemed the single most important idea in human history (Dennett, 1996). Somewhat paradoxically, it is also often branded a mere tautology, and when one thinks in terms of the survival of the fittest, there seems to be some basis for this view. However, considering the Darwinian scenario as a whole, it is easy to grasp its decidedly nontautological and nontrivial aspect. Indeed, what Darwin proposed is a mechanism for the transformation of random variation into adaptations that are not random at all, including elaborate, complex devices that perform highly specific functions and so increase the fitness of their carriers. Coached in physical terms and loosely following Erwin Schroedinger's famous treatise, Darwinian evolution is a machine for the creation of *negentropy*—in other words, order from disorder. I submit that this was the single key insight of Darwin, the realization that a simple mechanism, devoid of any teleological component, could plausibly account for the emergence, from random variation alone, of the amazing variety of life forms that appear to be so exquisitely adapted to their specific environments. Viewed from that perspective, the "invisible hand" of natural selection appears almost miraculously powerful, and one cannot help wondering whether it is actually sufficient to account for the history of life. This question has been repeatedly used as a rhetoric device by all kinds of creationists, but it also has been asked in earnest by evolutionary biologists. We shall see in the rest of this book that the answers widely differ, both between scientists and between different situations and stages in the evolution of life.

Of course, Darwinism in its original formulation faced problems more formidable and more immediate than the question of the sufficiency of natural selection: Darwin and his early followers had no sensible idea of the mechanisms of heredity and whether these mechanisms, once discovered, would be compatible with the Darwinian scenario. In that sense, the entire building of Darwin's concept was suspended in thin air. The rediscovery of genetics at the beginning of the twentieth century, followed by the development of theoretical and experimental population genetics, provided a solid foundation for Darwinian evolution. It was shown beyond reasonable doubt that populations evolved through a process in which Darwinian natural selection was a major component. The Modern Synthesis of evolutionary biology completed the work of Darwin by almost seamlessly unifying Darwinism with genetics. As it matured, Modern Synthesis notably "hardened" through indoctrinating gradualism, uniformitarianism, and, most important, the monopoly of natural selection as the only route of evolution. In Modern Synthesis, all changes that are fixed during evolution are considered adaptive, at least initially. For all its fundamental merits, Modern Synthesis is a rather dogmatic and woefully incomplete theory. To name three of the most glaring problems, Modern Synthesis makes a huge leap of faith by extending the mechanisms and patterns established for microevolution to macroevolutionary processes; it has nothing to say about evolution of microbes, which are the most abundant and diverse life forms on Earth; and it does not even attempt to address the origin of life.

Recommended further reading

Futuyma, Douglas. (2009) *Evolution*, 2d edition. Sunderland, MA: Sinauer Associates.

Probably the best available undergraduate text on evolutionary biology.

Gould, Stephen Jay. (2002) *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

The almost 1,500-page tome obviously is not for the feeble at heart, and not many will read it in its entirety. Nevertheless, at least the first part is valuable for its clear and witty presentation of the history of evolutionary biology and its pointed critique of Modern Synthesis.

Hartl, Daniel L., and Andrew G. Clark. (2006) *Principles of Population Genetics,* 4th edition. Sunderland, MA: Sinauer Associates.

An excellent, fairly advanced, but accessible textbook on population genetics.

Mayr, Ernst. (2002) *What Evolution Is*. New York: Basic Books.

A basic but clear and useful presentation of classical evolutionary biology by one of the architects of Modern Synthesis.

Schroedinger, Erwin. (1992) *What Is Life?: With "Mind and Matter" and "Autobiographical Sketches."* Cambridge, MA: Cambridge University Press.

The first edition of this wonderful book was published in 1944, on the basis of a series of lectures that Schroedinger (one of the founders of quantum mechanics) delivered in Edinburgh, where he stayed during World War II. Obviously outdated, but remarkably lucid, prescient, and still relevant in the discussion of the role of entropy and information in biology.

Index

Note: Page numbers followed by *n* are located in the endnotes.

a

b

C

d

e

h

i

m

n

o

p

q-r

S

t

U

V

W

x-y-z