PREFACE

Of all the claims on our curiosity, we want most to understand ourselves. What are we? How have we come to be what we are? What lies in our future? Many features of our lives depend on accidents of history. The time and place of our birth largely determine what language we first learn to speak, whether we are likely to be well fed and well educated, and receive adequate medical care. Many aspects of our future depend on events outside ourselves and beyond our control.

Yet, within us, there are constraints on our lives that brook relatively little argument. In some respects, we are at the mercy of our genomes. Under normal circumstances, all of our basic anatomy and physiology, and eye colour, height, intelligence, and basic personality traits, are ingrained in our DNA sequences. This is not to say that our genomes dictate our lives. Some constraints are tight—for instance, eye colour—but our genetic endowment also confers on us a remarkable robustness.

This robustness also is a product of evolution. Within the last century, lifestyles have changed with a rapidity hitherto unknown (except for the instants of asteroid impacts). We can meet and survive brutal stresses. Our talents have many opportunities to nurture themselves and to develop in novel ways. These are gifts of our genomic endowment: What genes control is the response of an organism to its environment.

The human genome is only one of the many complete genome sequences known. Taken together, genome sequences from organisms distributed widely among the branches of the tree of life give us a sense, only hinted at before, of the very great unity in detail of all life on Earth. This recognition has changed our perceptions, much as the first pictures of the Earth from space engendered a unified view of our planet.

Superimposed on this underlying unity is great variety. We ask: What is special about us? What do we share with our parents and siblings and how do we differ from them? What do we share with all other human beings and what makes us different from the other members of our species? What are the sources of our differences from our closest extant non-human relatives, the chimpanzees? What do we have in common and how do we diverge from other species of primates? Of mammals? Of vertebrates? Of eukaryotes? Of all other living things?

The complete sequences of human and other genomes reveal the underlying text of this story. We are beginning to understand how our lives shape themselves under the influence of our genes, epigenetic modifications, and our surroundings and life histories.

We are also beginning to be able to intervene. Genetic engineering of microorganisms is an established technique. Genetically-modified plants and animals exist, and are the subjects of lively debate. To override the genes for hair colour is trivial. Changes in lifestyle or behaviour can—to some extent—avoid or postpone development of diseases to which we are genetically at risk. Gene therapy offers the promise of rectifying some inborn defects. Novel gene editing techniques based on the CRISPR/Cas system offer revolutionary power in reshaping individuals, whole species, and ecosystems.

Fast, inexpensive sequencing has transformed genomics. The landmark goal, the $US1000 human genome, has been achieved. It is very likely that within the lifetime of many readers, human genome sequencing will be nearly universal. Already, hundreds of thousands of human individuals have had their full genomes sequenced, and many more are on the way. Many people have had sequences determined for individual genes. For example, mutations in *BRCA1*, *BRCA2*, and *PALB2* suggest an increased likelihood of developing breast or ovarian cancer. Many people have had these regions sequenced to provide information about their personal risk. This can be more than informative: some individuals—most famously the actress Angelina Jolie—have opted for prophylactic surgery.

The spectacular progress in high-throughput sequencing, and the resulting explosive growth of the data produced—in quality, quantity, and type—have altered the entire landscape of genomics itself, and its influence has invaded surrounding fields. No area of biology has been left unscathed.

Study of human disease through sequencing continues to be a major effort. The possibility of applications to improve human health were the major motivation for support of the Human Genome Project, and continue to be the major elicitor of largesse for pursuing its consequences. Identification of genes responsible for particular diseases permits testing, genetic counselling, and risk assessment and avoidance. Cancer genomics—the comparison of sequences from normal and tumour cells from single patients—has become a heavy industry.

Understanding the relationships between genes and disease will allow more precise diagnosis and warnings of increased risk of disease in patients and their offspring. For many important diseases, a patient can expect more precise diagnosis and prognosis, and more precise recommendations for treatment. Design of treatment based on a patient's genome, called pharmacogenomics, is already part of clinical practice. Genomes of other organisms also have implications for human health, especially those of pathogenic organisms that have developed, or are threatening to develop, antibiotic resistance. We study the biology of viruses and bacteria to take advantage of their vulnerabilities, and the biology of humans to ward off the consequences of our own.

This century will see a revolution in healthcare development and delivery. Walls between 'blue sky' research and clinical practice are tumbling down. It is possible that a reader of this book will discover a cure for a disease that would otherwise kill him or her. It is extremely likely that Szent-Györgyi's quip, 'Cancer supports more people than it kills', will come true. One hopes that this will happen because the research establishment succeeds in developing therapeutic or preventative measures against tumours, rather than by merely imitating their uncontrolled growth.

Other applications of genomics include improvement of crops and domesticated animals, enhancing food production, and support of conservation efforts dedicated to preserving endangered species. Development of alternative energy sources is a challenge to both physics and biology.

Beyond these applications, genomics offers us a profound understanding of fundamental principles of biology. The history of life is a pageant for which we are beginning to delineate the choreography. On the personal level, within this history, genome exegesis will fundamentally alter our perception of ourselves: what it means to be human.

In this book, I have tried to present a balanced view of the background of the subject, the technical developments that have so greatly increased the data flow, the current state of our knowledge and understanding of the data, and applications to medicine and other fields. With power derived from knowledge goes commitment to act wisely. We have responsibilities, to ourselves, to other people, to other species, and to ecosystems ranging up to the entire biosphere.

Ethical, legal, and social issues have been a prominent component of the human genome project. Most technical questions in genomics, as in other scientific subjects, have objectively correct answers. We do not know all the answers, but they are out there for us to discover. Ethical, legal, and social issues are different. Many choices are possible. Their selection is not the privilege of scientists in individual laboratories, but of society as a whole. Scientists do have a responsibility to contribute to the informed public discussion that is essential for wise decisions.

One aspect of the first edition that I liked was its concision. Unfortunately, this has had to be sacrificed to the stampeding progress of the field. A major problem encountered in writing about genomics is the need to pick and choose from the many riches of the subject. The list of subjects that cannot be left out is too long and threatens to reduce the treatment of each to superficiality. There is also a serious organizational challenge: many phenomena must be approached from several different points of view. A reader may be relieved to conclude that a topic has been beaten thoroughly into submission in one chapter, only to encounter it again, alive and kicking, in a later section, in a different context.

The speed at which the field is moving causes other problems. An author is often pleased with a draft of a section only to find the carefully described conclusions overturned in next week's journals. Yet, there is a great pleasure in seeing nature's secrets emerging before one's eyes.

Another casualty of rapid progress is the frequent dismissal of attention to history and biography. We are fantastically interested in the development of the sea urchin and the fruit fly, but not in the development of molecular genomics. This is too bad: those who do not learn from the successes of history will find

it harder to emulate them. Intellectual struggles that occupied entire careers leave behind only the terse conclusions, often without any appreciation of the experiments that established the facts, much less of the alternative hypotheses tested and rejected. We remember the breakthroughs, but not the frustrations that their achievement required. The force of the scientists' personalities, and their foibles, are forgotten. Making use of other people's results isn't the same thing as creating new ones: 'To imitate the *Iliad* is not to imitate Homer'.

Genomics is an interdisciplinary subject. The phenomena we want to explain are biological. But many fields contribute to the methods and the intellectual approaches that we bring to bear on the data. Physicists, mathematicians, computer scientists, engineers, chemists, clinical practitioners and researchers have all joined in the enterprise. This book will appeal, at least in part, to all of them. However, the central point of view remains focused on the biology.

More specifically, the focus is on human biology. In fact, on the biology of humans who are curious about other species, albeit primarily for what the other species tell us about ourselves. This choice naturally reflects the potential readership of this book. (If bacteria or fruit flies could read, genomics textbooks would look very different.)

In the new edition, extended coverage is given both to clinical applications to humans and to the applications of genome sequences to working out of evolutionary relationships in microorganisms, plants, and animals. Non-clinical applications to human biology and history are sufficient to justify a chapter of their own. The genomics of plant and animal domestications not only shed light on human—as well as plant and animal—history, but emphasize the reciprocal interactions between humans and the rest of the biosphere.

This book assumes that the reader already has some acquaintance with modern molecular biology, and builds on and develops this background, as a self-contained presentation. It is suitable as a textbook for undergraduates or starting postgraduate students.

Progress is happening so fast as to make unavoidable a feeling of frustration in aiming at a moving target. The hope is that the third edition has erected for the reader a sound framework, both intellectual and factual, that will make it possible, when encountering subsequent developments, to see where and how they fit in.

Exercises and problems at the end of each chapter, and 'weblems' on the Online Resource Centre, test and consolidate understanding, and provide opportunities to practise skills and explore additional subjects. Exercises are short and straightforward applications of material in the text. Answers to exercises appear on the website associated with the book. Problems, also, make use of no information not contained in the text, but require lengthier answers or, in some cases, calculations. The third category, 'weblems', requires access to the Internet. Weblems are designed to give readers practice with the tools required for further study and research in the field. Some of these are suitable for use as practical or laboratory assignments, or even as class projects.

Key terms are highlighted in the text, and are defined in the Glossary at the end of the book.