Preface

"Nothing in biology makes sense except in the light of evolution."

Theodosius Dobzhansky

It is very nearly 150 years since Charles Darwin published the *Origin of Species* (1859). Very deliberately, Darwin scarcely mentioned human evolution in the 'Origin' and left Thomas Huxley, in his *Evidence as to Man's Place in Nature* (1863), to present anthropological and anatomical evidence to support his contention that humans had evolved from other primates. Later, however, in his *Descent of Man* (1871), Darwin opined that 'man was the co-descendant with other mammals of a common progenitor' and hence 'still bears in his bodily frame the indelible stamp of his lowly origin'.

It is sometimes forgotten that evolutionary ideas preceded Charles Darwin. Indeed, while Darwin's own grandfather Erasmus, in his *Zoonomia* (1796), had anticipated Jean-Baptiste de Lamarck's ideas on the inheritance of acquired characteristics, he also stimulated the thinking of his future grandson. In 1798, Thomas Malthus published his *Essay on the Principle of Population* which was to have such an important influence on evolutionary thinking. In an especially prescient essay published in 1818 (*An Account of a Female of the White Race of Mankind, Part of Whose Skin Resembles That of a Negro*), a Scottish-American physician, William Charles Wells, appears to have anticipated subsequent ideas of natural selection by speculating that the 'dark races of mankind' might be better adapted to warmer climes than the 'white races' and hence might enjoy an immunity from certain tropical diseases that would allow them to 'become the most prevalent, if not the only race, in the particular country in which [they] had originated'. It is of course only right and proper here to make mention of Alfred Russell Wallace whose *On the Law Which has Regulated the Introduction of New Species* (1855) and *On the Tendency of Varieties to Depart Indefinitely from the Original Type* (1858) finally prompted Darwin to publish his *Origin of Species* in 1859. In passing, it should also be noted that in his *Hereditary Genius* (1869) and *Natural Inheritance* (1894), Darwin's half cousin Francis Galton described pioneering work on the heritability of human variation, particularly in the sphere of ability, intelligence and stature.

Lamarck's ideas on the inheritance of acquired characteristics were finally to be superseded once August Weismann had articulated his germ-plasm theory of heredity (1893). The subsequent rediscovery of Mendel's laws of heredity by Carl Correns, Hugo de Vries and Erich von Tschermak-Seysenegg in 1900, Walter Sutton's (1902) formulation of the chromosomal theory of heredity, von Dungern and Hirszfeld's (1910) demonstration that human blood groups are inherited, T.H. Morgan's *Drosophila* studies (1908–1930) which provided experimental verification that genes reside on chromosomes, the early studies of William Bateson (1890–1910), Hugo de Vries (1890–1905) and Hermann Muller (1920s) on mutation, all helped to set the scene for J.B.S. Haldane, R.A. Fisher and Sewall Wright to develop a mathematical basis for population genetics and hence to effect a synthesis of Darwinism and Mendelism that has served as the foundation of the theory of evolution as we know it today. The 'neutral theory of molecular evolution', which was put forward in the late 1960s by Motoo Kimura, Thomas Jukes and others to underpin the concept of 'molecular clocks', sparked a lively selectionist–neutralist debate, but is not incompatible with adaptive evolution, and nowadays productively inhabits the same intellectual landscape.

In the 1950s and 1960s, the pioneering studies of Arthur Mourant and Henry Harris on, respectively, human blood group and protein polymorphism provided a glimpse of the likely extent of human genetic variability. In 1954, Anthony Allison reported that the prevalence of sickle cell anaemia carriers in East Africa correlated with the incidence of malaria, consistent with the postulate that selection was favouring sickle cell heterozygotes on the basis of their relative resistance to the disease. By the 1960s, Luca Cavalli-Sforza and Anthony Edwards were using protein polymorphism data to construct phylogenies to explore how genes had become differentially distributed between human populations by migration and admixture. With the elucidation of the structure of deoxyribonucleic acid (DNA) (1953) and the subsequent development in the early 1970s of recombinant DNA (molecular cloning) and DNA sequencing techniques, evolutionary biology entered the modern era, allowing the 'genocentric view' of evolution to move to centre stage. In the process, protein markers have now become largely superseded by DNA polymorphisms and it is the latter (particularly mitochondrial and Y-chromosome markers) that have proved extremely useful in the reconstruction of the evolutionary history of human populations and for the study of the emergence of modern humans out of Africa. With the advent of PCR (the *polymerase chain reaction*: a fast, inexpensive technique for making an unlimited number of copies of any piece of DNA) in 1985, an innovation that made automated low-cost large-scale DNA sequencing a reality, the whole nuclear genome has now become amenable to detailed population genetic analysis.

To appreciate fully the relationship between the structure and function of our genes and the proteins they encode, we must be aware of their evolutionary past and their molecular ontogeny. The revolution in human molecular genetics that has taken place over the last three decades has yielded a wealth of information on the structure and function of genes, gene expression, mutation, polymorphic variation and linkage disequilibrium.

Furthermore, as sequencing efforts were scaled up through automation during the 1990s, the focus moved from genes to genomes. Determination of the complete genome sequences of yeast (1996), nematode worm (1998) and fruitfly (2000) were followed in short order by human (2001–2003), mouse (2002), pufferfish – *Fugu* (2002), sea squirt (2002), rat (2004), chicken (2004), pufferfish – *Tetraodon* (2004), dog (2005), chimpanzee (2005), sea urchin (2006), sea anemone (2007), macaque (2007), opossum (2007), cat (2007), platypus (2008) and amphioxus (2008). Even though the annotation of our full complement of approximately 30 000 genes is still in progress, genome-wide studies are already yielding abundant evidence for the signatures of past selection/adaptive evolution within human gene sequences. The completion of the sequencing of the three billion base pair human genome, coupled with the availability of an ever increasing number of other vertebrate and invertebrate 'model genomes', has ushered in a new era of *comparative genomics*. Indeed, comparative genomics is proving to be an invaluable aid in a variety of different contexts, from helping us to probe the functions of noncoding sequences in the human genome to identifying the mutational changes that have given rise to the veritable abundance of biochemical and morphological variation that characterizes extant organisms. We are thus at last in a position to be able to study with relative ease the fine structure of the molecules that Zuckerkandl and Pauling (1965) presciently described as being 'documents of evolutionary history', to reconstruct evolutionary lineages not only at the level of the gene and the chromosome but also, most powerfully of all, at the level of the genome.

The sequencing of the chimpanzee genome and its subsequent comparison with its human counterpart, have begun to reveal the spectrum of genetic changes accompanying human evolution. Segmental duplication is now known to have been a frequent event during the evolution of the human genome. In addition, numerous human-specific polymorphic gains and losses of genes (copy number variation) have been identified as well as changes in gene expression. The scale of the structural diversity (polymorphism) evident within both lineages has, however, hampered the analysis of the structural divergence between the human and chimpanzee genomes. The concomitant evaluation of genetic divergence and diversity at the nucleotide level is nevertheless leading to the identification of many genes that have evolved under positive selection and which may thus have been involved in the development of human lineage-specific traits. We are thus acquiring the ability to ask searching questions about our origins, about the demographic processes associated with the global radiation of humankind, as well as some of the unique adaptations that make us human. Knowledge of our origins, and of our relationship to the rest of the natural world, has the potential to enrich a wide range of human activities. Indeed, the study of human evolution is arguably fundamental to our continuing quest for self-knowledge. As McConkey and Goodman (1997) opined:

> *It is our conviction that comparative analysis of human and ape genomes is far more than an excursion into natural history at the molecular level. Until we have a detailed understanding of the genetic differences between ourselves and our closest evolutionary relatives, we cannot really know what we are.*

Both the gene and the genome have been dynamic entities over evolutionary time. This dynamism manifests as genetic variation that is readily apparent not only between related species but also within a species. Thus, DNA copy number variation represents a very considerable source of human genetic diversity. Indeed, two human genomes may differ by more than 20 Mb of their 3300 Mb (involving potentially a couple of thousand genes) and it is likely that the full extent of inter-individual variation still remains to be discovered. This high degree of variability in gene copy number between any two individuals must surely serve to challenge definitions of 'normality'. Many genes have, over the millennia, undergone gross rearrangement as a result of the action of any one of a number of mutational processes such as insertion, inversion, duplication, repeat expansion, translocation or deletion. It turns out that even relatively conserved genes do not necessarily always change by a slow incremental process of single base-pair substitution. Indeed, such genes may acquire multiple DNA sequence changes simultaneously by mechanisms such as gene conversion. Elucidation of the consequences of these mutational changes for the phenotype of the organism represents a formidable challenge for biologists. However, new fields of enquiry such as transcriptomics and proteomics promise to provide windows onto the evolution of gene expression and its downstream transcriptional and translational complexity.

Evolutionary studies are proving of great importance to molecular medicine, from the identification of new disease genes to understanding how specific gene lesions exert their pathological effects. Indeed, evolutionary principles are not only helping geneticists to understand population differences in the frequency of inherited diseases but they are also bringing about vital improvements in the design of disease association studies. Evolutionary principles additionally find numerous applications in the context of understanding cancer and infectious disease. Ernst Mayr once said that 'there is not a single *Why?* question in biology that can be answered adequately without a consideration of evolution'. It is now becoming clear that the traditional mechanistic *How?* questions in medicine are increasingly metamorphosing into *Why?* questions whose answers must ultimately reside in the realm of evolutionary explanation. Meanwhile, the unifying principles of evolution are helping to make sense of genetic, biochemical, epigenetic and morphological variation in a human context, thereby providing the conceptual framework which is a *sine qua non* for understanding function and dysfunction in medicine. Appreciating that our bodies have evolved, both structurally and mechanistically, in ways that are not necessarily conducive to our long-term health and wellbeing, promises to be a major step on the road

towards the development of new therapeutic approaches to common disease.

At the heart of the way in which we conceptualize the process of evolution is an apparent dichotomy: on the one hand, sequence conservation is usually held to imply functionality, on the other, the emergence of novel functions implies change. Thus, selection may act conservatively so as to retain features of structural or functional importance (negative or purifying selection), or act so as to favour changes that confer some advantageous characteristic (positive selection). As mentioned earlier, evolution can also proceed in a neutralist fashion, some features being acquired not necessarily because of any selective advantage accruing to the organism but rather owing to the vagaries of population size, structure or dynamics over an extended period of time. It has also become clear that some DNA sequences either possess or acquire a mutational momentum of their own and, in the absence of negative selection, can drive the process of evolutionary change without the necessary involvement of positive selection. In practice, these diverse mechanisms have acted in different combinations and permutations at many loci simultaneously. Evolution has no foresight, however, and the ultimate arbiter of evolutionary success is always the number of surviving and fertile offspring.

Evolutionary biology has now become so broad that its impact may be felt across the spectrum of the biological sciences. Indeed, as P.W. Ewald has put it, 'Evolutionary biology is so firmly integrated with the rest of biology that it is not possible to mark a boundary between them'. In the process, the study of evolution has been revolutionized, with the emphasis being shifted to some extent away from the traditional domains of anthropology, archaeology, palaeontology, comparative anatomy and physiology, morphology and demography, and towards the interpretation of the molecular evidence, whether from extant organisms, reconstructed ancestral genomes or, as in the case of Neanderthals, from the bones of extinct hominids. While this has provided us with an unparalleled opportunity to identify and track the veritable multitude of molecular events (from the chromosomal level down to the single base-pair) that have occurred during vertebrate, mammalian, primate and hominid evolution, we should not lose sight of the importance of the more traditional lines of enquiry. As new areas of scientific enquiry are spawned, the challenge will be to integrate information from an increasingly disparate series of specialties whose practitioners inevitably come to share less and less common ground. The *Handbook of Human Molecular Evolution* attempts to counter the trend towards fragmentation by promoting an integrated approach to the study of human evolution and emphasizing the interplay between molecular genetic concepts and principles on the one hand, and information acquisition and interpretation on the other. It is hoped that the *Handbook* will encourage researchers not to lose sight of the 'big picture', while also helping to translate the often arcane advances of specialists into language that is intelligible to the widest possible audience.

The aim of the *Handbook of Human Molecular Evolution* is relatively straightforward: to bring together under the same cover the many and varied strands of our knowledge of human/primate/vertebrate molecular evolution. Hence, the 282 chapters that comprise this work have been arranged into 12 sections: General Concepts in Evolutionary Genetics; Mutation, Adaptation and Natural Selection; Evolutionary and Population Genetics; Human Evolution; Human Genome Evolution; Evolution of Human Gene Structure and Function; Evolution of Gene Expression; Mitochondrial Genome Evolution; Chromosomal Evolution; Comparative Genomics; Evolution and Disease Susceptibility; Analysis of Ancient DNA. This conceptual outline informed the selection of the chapters themselves and the connections between them. Some of these chapters are intended to be introductory, some are overviews that address topics of broad interest and importance, whereas others focus on quite specialized topics. The chapters are richly supplied with website information to allow access to relevant data sources over the internet. In any work of this size and scope, some overlap between chapters is unavoidable, but the Editors have taken the view that some degree of 'friendly redundancy' is actually desirable. All chapters are available online via the *Encyclopedia of Life Sciences* (http://www.els.net).

We are most grateful to the multitude of authors from more than 30 countries, who have given generously of their time to share their expertise, and to Andrea Baier, Martine Bernardes-Silva, Daniel Finch, Fiona Hutton, Liz Riordan and Louise Rush of John Wiley & Sons Ltd, without whose enthusiasm and dedication this *Handbook* would not have been possible.

David N. Cooper and Hildegard Kehrer-Sawatzki
Cardiff and Ulm
August 2008

How to use this book

Articles are divided into three different categories indicating their level of complexity: Introductory, Advanced and Keynote. Introductory articles have been written primarily for undergraduates and non-specialists requiring the basic concepts of a particular subject. Advanced articles provide a more detailed discussion of specialist subjects, equivalent to that found in graduate level texts. Keynote articles provide a platform for debate where controversial issues and 'hot topics' can be discussed.