GS
Garland Science

HUMAN EVOLUTIONARY GENETICS

second edition

Jobling · Hollox · Hurles · Kivisild · Tyler-Smith

This page intentionally left blank
to match pagination of print book

HUMAN EVOLUTIONARY GENETICS

second edition

HUMAN EVOLUTIONARY GENETICS

second edition

Mark Jobling

Edward Hollox

Matthew Hurles

Toomas Kivisild

Chris Tyler-Smith

Garland Science
Taylor & Francis Group
NEW YORK AND LONDON

Garland Science

Vice President: Denise Schanck
Senior Editor: Elizabeth Owen
Assistant Editor: Dave Borrowdale
Production Editor: Georgina Lucas
Illustrator: Matthew McClements, Blink Studio Ltd.
Cover Design: Matthew McClements, Blink Studio Ltd.
Copyeditor: Sally Huish
Typesetting: EJ Publishing Services
Proofreader: Jo Clayton
Indexer: Medical Indexing Ltd.

Front Cover

Antony Gormley
FEELING MATERIAL XIV, 2005
4 mm square section mild steel bar
225 x 218 x 170 cm (unextended size). 70 Kg
Photograph by Stephen White, London
© the artist

© 2014 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. Every effort has been made to trace copyright holders and to obtain their permission for the use of copyright material. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. No part of this book covered by the copyright herin may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

ISBN: 978-0-8153-4148-2

Library of Congress Cataloging-in-Publication Data

Human evolutionary genetics / Mark Jobling ... [et al.]. -- 2nd ed.
p. ; cm.
Rev. ed. of: Human evolutionary genetics / Mark A. Jobling, Matthew Hurles, Chris Tyler-Smith. c2004.
Includes bibliographical references and index.
ISBN 978-0-8153-4148-2
I. Jobling, Mark A. II. Jobling, Mark A. b Human evolutionary genetics.
[DNLM: 1. Evolution, Molecular. 2. Adaptation, Biological. 3. Anthropology, Physical. 4. Genetic Variation. 5. Genome, Human. QU 475]
QH431
599.93'5--dc23

2013008586

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 711 Third Avenue, New York, NY 10017, USA, and 3 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Garland Science
Taylor & Francis Group

Visit our web site at http://www.garlandscience.com

The authors

Professor Mark A. Jobling
Department of Genetics
University of Leicester, UK

Dr. Edward Hollox
Department of Genetics
University of Leicester, UK

Dr. Matthew Hurles
Wellcome Trust Sanger Institute
Hinxton, UK

Dr.Toomas Kivisild
Division of Biological Anthropology
University of Cambridge, UK

Dr. Chris Tyler-Smith
Wellcome Trust Sanger Institute
Hinxton, UK

PREFACE

This book is a completely revised edition of *Human Evolutionary Genetics*, first published in 2004. We decided to write the first edition because there were no textbooks available covering the areas that interested us. Once we had embarked upon the Herculean task of producing it, we realized why nobody had attempted to summarize this forbiddingly broad and contentious field before. But luckily the reception was positive, one eager person (our ideal reader and not, we point out, one of the authors in disguise) writing on Amazon that "I bought a copy for myself, and another one for my advisor. I have read it twice in a week!" A revised version seemed like a pretty good idea.

We cheerfully imagined that the second edition would be easier to write than the first. How wrong we were. First, all three original authors (MJ, MH, and CTS) had accumulated additional responsibilities that reduced the available time for writing. Second, the field obstinately continued to grow, and scarcely a week went by without some interesting and important development—the genomes of new species, genomewide surveys of human variation, next-generation sequencing and its data tsunami, spectacular ancient DNA discoveries, large-scale population studies, novel statistical methods, archaeological and paleontological revelations—the list goes on. We sometimes wished everyone would just stop working for a bit, so we could catch up. So, our deadline for the second edition passed, was revised, and passed again. HEG1 was becoming more and more out of date. We needed help.

The cavalry duly arrived in the form of two sterling new recruits to the authorial team—EH and TK. They brought their own areas of interest and expertise, but also a more efficient and energetic approach to the writing process, which revitalized the whole project. So, after a lengthy and difficult gestation, here is HEG2.

Following an initial introductory chapter, the book is divided into five sections, allowing it to be read by interested students and researchers from a broad range of backgrounds. "How do we study genome diversity?" (Chapters 2–4) and "How do we interpret genetic variation?" (Chapters 5–6) together provide the necessary tools to understand the rest of the book. The first of these sections surveys the structure of the genome, different sources of genomic variation, and the methods for assaying diversity experimentally. The second introduces the evolutionary concepts and analytical tools that are used to interpret this diversity. The subsequent two sections take an approximately chronological course through the aspects of our current state of knowledge about human origins that we consider most important. The section "Where and when did humans originate?" (Chapters 7–9) first considers our links to our closest living nonhuman relatives, the other great apes, then investigates the genetic changes that have made us

human, and finally details the more recent African origin of our own species. "How did humans colonize the world?" (Chapters 10–14) describes how human genetic diversity is currently distributed globally and then discusses the evidence for early human movements out of Africa, and the subsequent processes of expansion, migration, and mixing that have shaped patterns of diversity in our genomes. Finally, "How is an evolutionary perspective useful?" (Chapters 15–18) demonstrates the wider applications of an evolutionary approach for our understanding of phenotypic variation, the genetics of diseases both simple and complex, and the identification of individuals. Extensive cross-referencing between these sections facilitates different routes through the book for readers with divergent interests and varying amounts of background knowledge.

An important feature is the use of "Opinion Boxes"—short contributions by guest authors who are experts in different aspects of this diverse subject area. These help to give a flavor of scientific enquiry as an ongoing process, rather than a linear accumulation of facts, and encourage the reader to regard the published literature with a more critical eye. Opinions about how data should be interpreted change, and often an objective way to choose between different interpretations is not obvious. This is particularly true of genetic data on human diversity. Many of the debates represented in the Opinion Boxes scattered through this book derive from methodological differences.

Additional resources have been incorporated to permit interested readers to explore topics in greater depth. Each chapter is followed by a detailed bibliography, within which the sources that should be turned to first for more detail are highlighted in purple text. Electronic references to internet sites are given throughout the book, both for additional information and for useful software and databases. We explain specialist terms where they are first used, and include an extensive glossary at the back of the book that defines all terms in the text that are in bold type. At the end of each chapter is a list of questions (some short-answer, and some prose) that allow the reader to test their knowledge as they proceed. Teachers may be interested to know that most of the figures are freely available from the Garland Science Website (www.garlandscience.com) for use in teaching materials.

An obvious difference from the first edition is the presence of two extra chapters, reflecting developments in understanding the human genome in the context of other hominid genomes, and in complex disease. A very welcome development is the availability of full-color printing, which makes complex figures much easier to understand.

ACKNOWLEDGMENTS

We have many people to thank for contributions to this book. Twenty-five researchers found time to write Opinion Boxes; we are very grateful to: Mark Achtman, Elizabeth Cirulli & David Goldstein, Graham Coop & Molly Przeworski, Dorian Fuller, Tom Gilbert, Boon-Peng Hoh & Maude Phipps, Doron Lancet & Tsviya Olender, Daniel MacArthur, Andrea Manica & Anders Eriksson, Linda Marchant & Bill McGrew, John Novembre, Bert Roberts, Aylwyn Scally & Richard Durbin, Sarah Tishkoff, George van Driem, Bernard Wood, and Richard Wrangham & Rachel Carmody. We must stress that any opinions outside these Boxes are our own, and not necessarily endorsed by the Opinion Box contributors.

We are very grateful to the external specialists who helped improve the manuscript by providing data and advice, commenting on figures and questions, and reviewing draft chapters. Of course, despite the efforts of reviewers, errors and omissions will remain, and we take full responsibility for these. Thanks to: John Armour (University of Nottingham), Chiara Batini (University of Leicester), Stefano Benazzi (University of Vienna), Antonio Brehm (University of Madeira), John Brookfield (University of Nottingham), Terry Brown (University of Manchester), Anne Buchanan (Pennsylvania State University), John M. Butler (The National Institute of Standards and Technology), Lucia Carbone (Oregon Health and Science University), Susana Carvalho (University of Oxford), Vincenza Colonna (Wellcome Trust Sanger Institute), Murray Cox (Massey University), Todd R. Disotell (New York University), Michael Dunn (Max Planck Institute for Psycholinguistics), Wolfgang Enard (Max Planck Institute for Evolutionary Anthropology), Greg Gibson (Georgia Institute of Technology), Tom Gilbert (Natural History Museum of Denmark), Bernard Grandchamp (University of Paris Diderot), Ryan Gutenkunst (University of Arizona), Wolfgang Haak (University of Adelaide), Phillip Habgood (University of Queensland), Pille Hallast (University of Leicester), Terry Harrison (New York University), Simon Hay (University of Oxford), Andy I.R. Herries (La Trobe University), Sarah Hill (University of Cambridge), Hirohisa Hirai (Kyoto University), Rosalind Howes (University of Oxford), Arati Iyengar (University of Central Lancashire), Peter de Knijff (Leiden University Medical Center), Vincent Macaulay (University of Glasgow), Ripan Malhi (University of Illinois Urbana-Champaign), Nicola Man (University of New South Wales), Tomas Marques-Bonet (University Pompeu Fabra), Celia May (University of Leicester), Patrick McGrath (Georgia Tech), Bill McGrew (University of Cambridge), Pierpaolo Maisano Delser (University of Leicester), Mait Metspalu (University of Tartu), Darren Monckton (University of Glasgow), Maru Mormina (University of East Anglia), Connie Mulligan (University of Florida), David Nelson (Baylor College of Medicine), Barbara Ottolini (University of Leicester), Svante Pääbo (Max

Planck Institute for Evolutionary Anthropology), Luca Pagani (Wellcome Trust Sanger Institute), Luisa Pereira (University of Porto), Fred Piel (University of Oxford), Sohini Ramachandran (Brown University), Christine Rees (Illumina), Tim Reynolds (Birkbeck, University of London), Jorge Macedo Rocha (University of Porto), Rebecca Rogers Ackermann (University of Cape Town), Alexandra Rosa (University of Madeira), Mark Seielstad (University of California San Francisco), Giorgio Sirugo (Tor Vergata University Medical School), Roscoe Stanyon (University of Florence), Jay Stock (University of Cambridge), Peter Sudbery (University of Sheffield), Dallas Swallow (University College London), Martin Tobin (University of Leicester), Richard Villems (University of Tartu), Cynthia Vigueira (Washington University St. Louis), Bence Viola (Max Planck Institute for Evolutionary Anthropology), Tim White (University of California, Berkeley), Mike Whitlock (University of British Columbia), Jinchuan Xing (Rutgers University), Yali Xue (Wellcome Trust Sanger Institute), and Bryndis Yngvadottir (Wellcome Trust Sanger Institute).

We thank Liz Owen at Garland for her persistence, pragmatism, and practical help in bringing HEG2 to fruition. Dave Borrowdale's efficient administration of the review process of text and figures has also been greatly appreciated. We thank the artist, Matt McClements, for the professional job he has done in redrawing our figures. We also thank Georgina Lucas for helping us through the proof stage.

We are also grateful for the funding bodies and institutions that have allowed us to maintain our research interests in human evolutionary genetics, and to produce this book. In particular, MJ thanks the Wellcome Trust for funding through a Senior Fellowship, and the University of Leicester for providing a collegial working environment. EH thanks the same institution for giving him gainful employment. TK thanks colleagues from the Division of Biological Anthropology, University of Cambridge, for their support; and MH and CTS thank the Wellcome Trust both for its direct support and for its creation of the stimulating environment of the Sanger Institute.

Writing a textbook is undoubtedly interesting and educational, but also onerous. Much of this burden has fallen on our families and colleagues, as well as ourselves. We all thank our current and former group members for putting up with the distractions of book preparation, and for interesting delicacies from around the world. In addition, MJ thanks Nicky, Bill, and Isobel, EH thanks Gill and Kirsten, MH thanks Liz, Edward, Jenny, and Audrey, TK thanks Dagne and Uku, and CTS thanks Yali and Jack. Perhaps this book can now provide an explanation for our preoccupations and absences.

CONTENTS

DETAILED CONTENTS

AN INTRODUCTION TO HUMAN EVOLUTIONARY GENETICS

CHAPTER ONE

In this chapter, we introduce human evolutionary genetics and describe the recent dramatic advances in knowledge and understanding that have led to its central role in all human genetic studies. We explore the diverse sources of information about the human past that are available to us. These complementary records tell us about different aspects of the past, and are informative over different time-scales. We also identify the fundamental human evolutionary questions that can now be addressed using genetics.

1.1 WHAT IS HUMAN EVOLUTIONARY GENETICS?

Evolutionary genetics is founded on the principle that the genetic record of life is contained in the **genomes** of living species and it reveals evolutionary processes and relationships all the way back to the last universal common ancestor of all species. To find out about this ancestral organism we have to compare and contrast the most distantly related branches on the tree of life. Comparisons among much more closely related individuals, such as those from the same species, provide evidence on much more recent evolutionary processes. Our ability to read this genetic record has developed enormously in the last few years, although our confidence that information on our past exists within our heritable material is somewhat older. Genetic evidence comes from two main sources:

- The genomes of living individuals that must have been passed down from ancestors
- Ancient DNA from well-preserved organic remains, which may or may not be represented now in living descendants

Human evolutionary genetics involves the study of how different copies of the human genome differ from one another, and from those of our closest relatives, other **primates**. Differences between genomes also form the basis of anthropological, medical, and **forensic** genetics. All of these fields are experiencing massive advances as a result of two developments. First, the public availability of human and nonhuman genome sequences, annotated with important functional elements such as **genes**, and with sites of genetic variation. Second, technology allowing analysis of most of this genetic variation across the genome, initially with the development of hybridization **microarrays** (or **chips**), followed by the development of methods to sequence whole genomes rapidly and cheaply (so-called **next-generation sequencing**). This explosion of information is driving an unprecedented period of innovation that gives us the tools to analyze huge datasets of unparalleled quality and quantity. This wealth of data is itself catalyzing the development of new interpretative methods. In addition, with the publication of genome sequences from other

extant primates, as well as data from the genomes of extinct members of the genus *Homo*, we can now comprehensively catalog the genetic differences between humans and our closest relatives. As we shall see in this book, these impressive developments are allowing us to answer some of the most fundamental questions regarding human origins. These are indeed exciting times for human evolutionary genetics.

In times past, when writing materials were precious commodities, a scribe would often reuse an existing manuscript rather than obtain a new parchment. The manuscript would be turned through 90°, and overwritten. These overwritten manuscripts bearing the imprint of more than one text are known as **palimpsests**. The genetic record is similarly a complex palimpsest. Variation among modern individuals is shaped by cumulative past processes, which allows us to investigate different times in human prehistory with the same data.

Different layers of the past are accessible through the analysis of genetic diversity. Moving from the most ancient to the most recent, we encounter:

- Our phylogenetic relationship to other species (Chapter 7)
- The origin of our species (Chapters 8, 9)
- Prehistorical migrations (Chapters 10, 11, 12, and 13)
- Historical migrations (Chapter 14)
- Genealogical studies (Chapter 18)
- Paternity testing (Chapter 18)
- Individual identification (Chapter 18)

Extracting information on any one past period or event requires careful interpretation to isolate it from previous and subsequent processes. This information can tell us not only about the **demographic history** and origins of populations but also something about the environmental challenges faced by those populations, through the influence of **natural selection** on genetic variation.

It is said that "the past is the source of the present" and this is true in the academic field of human evolutionary genetics as much as elsewhere. This exciting subject owes its current status to developments and debates over the last 150 years in genetics, paleontology, archaeology, anthropology, and linguistics. In this book we have avoided cataloging this history, instead taking a twenty-first-century perspective, but we discuss key developments where they are relevant, and provide a time line in **Figure 1.1**.

1.2 INSIGHTS INTO PHENOTYPES AND DISEASES

What use can an evolutionary perspective on human genetic variation have beyond the reconstruction of the past for its own sake?

A shared evolutionary history underpins our understanding of biology

The great twentieth-century evolutionary biologist Theodosius Dobzhansky wrote that "Nothing in biology makes sense except in the light of evolution." All the sizes, shapes, chemistries, and genes of organisms alive today derive from ancestors that can be traced back over billions of years. All of these features have been shaped by the environmental challenges faced by these organisms and their ancestors. If it were not the case that humans share an ancestor with every other species on the planet, there would be no value in performing any form of comparative analysis. There would be nothing that the *Escherichia coli* bacterium, brewer's yeast, fruit fly, nematode worm, zebrafish, mouse, or **chimpanzee** could tell us about ourselves. None of these species is our ancestor: they are our cousins, equally distant in time from our common ancestor (**Figure 1.2**). It is our shared evolutionary heritage with these species that makes them such powerful **model organisms**.

1786	Recognition of language families
1856	Discovery of Neanderthal type specimen
1859	Publication of Darwin's "The Origin of Species"
1866	Publication of Mendel's "Experiments in Plant Hybrids"
1871	Publication of Darwin's "The Descent of Man"
1900	Discovery of first genetic polymorphism—ABO blood group (Landsteiner)
1908	Hardy–Weinberg principle formulated
1918	Fisher reconciles Darwin's natural selection and Mendel's mechanism of inheritance
1925	*Australopithecus* fossil described from South Africa
1930–32	Fisher, Haldane & Wright publish the foundations of modern population genetics
1944	DNA shown to be heritable material
1949	Radiocarbon dating introduced
1953	Double-helical structure of DNA described
1956	Human chromosome number described
1957	Hemoglobin amino acid sequences determined
1959	Y chromosome shown to be sex-determining
1966	Genetic code deciphered
1968	Neutral theory of molecular evolution (Kimura)
1969	Internet first successfully tested
1977	Publication of DNA sequencing methods
1978	First human restriction fragment length polymorphisms (RFLPs) described
1978	First human *in vitro* fertilization
1980	First genome (φX174 bacteriophage) sequenced
1981	Human mitochondrial DNA (mtDNA) genome sequenced
1984	DNA fingerprinting (minisatellites) discovered
1984	DNA-DNA hybridization shows human–chimpanzee common ancestry
1985	Invention of polymerase chain reaction (PCR)
1985	First human ancient DNA results published
1985	First Y-chromosomal polymorphism described
1987	Development of laser-induced fluorescent detection of DNA
1987	African origin of human mtDNA identified
1988	Launch of Human Genome Project
1989	Development of capillary electrophoresis for sequencing
1990	First human microsatellites described
1991	Human Genome Diversity Project proposed
1994	Publication of "The History and Geography of Human Genes" (Cavalli-Sforza et al.)
1996	First mammal cloned from adult cell (Dolly)
1997	First Neanderthal mtDNA sequence
1999	First human chromosome sequenced (Chr 22)
2001	Release of draft human genome sequence
2002	Release of draft mouse and *Plasmodium* genome sequences
2002	Human Genome Diversity Project (HGDP) Cell Line Panel released
2004	First maps of copy-number variation published
2005	First-generation human Haplotype Map (HapMap) published
2005	Release of draft chimpanzee genome sequence
2005	First development of next-generation sequencing methods
2006	1 Mb of Neanderthal genomic sequence published
2007	First large-scale genomewide association studies
2007	First personal human genome resequenced (Venter)
2007	Second-generation human Haplotype Map (HapMap) published
2009	Exome capture and sequencing methods published
2010	Denisovan mtDNA and genome sequences published
2010	1000 Genomes Project pilot study published
2012	All great ape genomes now sequenced

Figure 1.1: Time line of important developments in the field of human evolutionary genetics.

To take just one example, sequencing the mouse genome allows us to identify more genes in the human genome than does sequencing the human genome alone. By identifying segments of DNA that are more similar between the two species than could be expected by chance, we can identify regions whose evolution has been constrained by the need to perform a specific function. Some of these regions are genes. In other words, we can identify a gene not because it

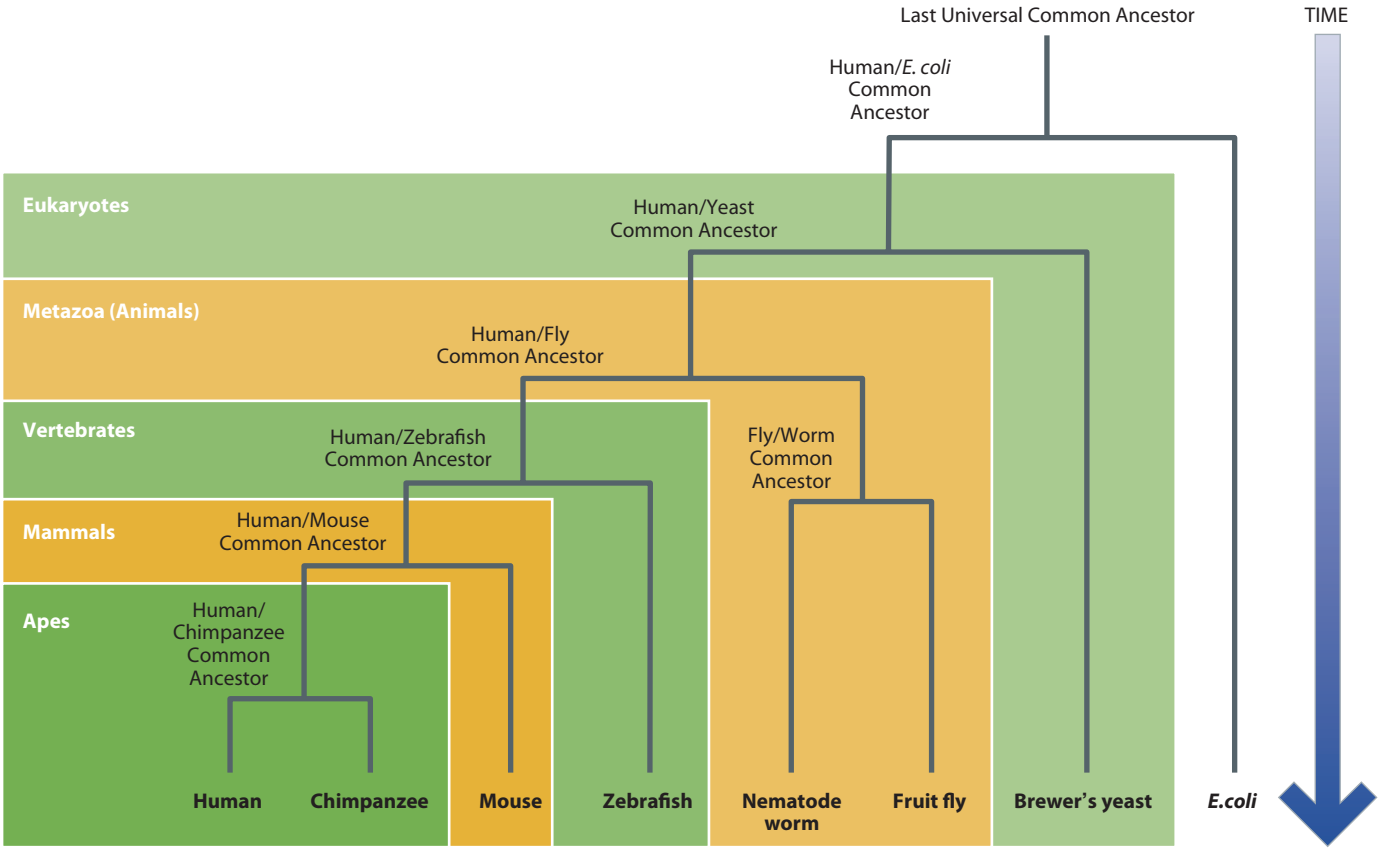

Figure 1.2: Cousins, not ancestors. A **phylogenetic tree** relates different branches of modern species, showing that they are all equally derived from their common ancestors in terms of time. Here branch lengths reflect evolutionary time, rather than evolutionary change.

looks like a gene, nor because an organism treats it like a gene (that is, makes a product from it), but because it *evolves* like a gene.

Understanding evolutionary history is essential to understanding human biology today

If we were to take a perspective to the biology of modern humans that neglected evolutionary history, what might we predict about the genetic diversity of our species, the significance of our phenotypic differences, and the prevalence of disease-causing **alleles**?

First, we would be struck by the huge numbers of humans, especially when compared with other animals of similar size. We might reasonably think that this should be mirrored by a correspondingly greater genetic diversity. Second, we might note the clustered distribution of phenotypic diversity (for example, skin color) among modern human groups and might expect this to be matched by a similar structuring of genetic diversity. Third, we might suppose that disease-causing alleles would be specific to different continental groups, in a similar manner to some of their easily observable "normal" phenotypes. As we shall discover in this book, all of these conclusions would be wrong.

To understand why this is so, we must comprehend that the past is not simply something that happened, and is studied for its own sake, but is more properly considered as the source of the present. If we are to improve our present circumstances, we must take account of how that present has come to be. An evolutionary perspective does not just address the question, What happened in the past? but also the question, Why is the present like it is?

Once we understand that the obvious differences between peoples' appearances can be unreliable indicators of biological origins, we start to appreciate the other factors that have shaped and continue to shape human biology. The interaction of humans and their surroundings comes to the fore, as does an

understanding of human adaptability in the face of huge variability in inhabited environments.

Understanding evolutionary history shapes our expectations about the future

An evolutionary perspective on human genetic variation also allows us to make predictions, both about biological research, and about the future of our species. Today, we can pose many more biological questions than we are able to answer. An evolutionary perspective tells us how we might go about answering these questions, and about what kinds of answers we might expect.

Phenotypic traits of humans, be they skin color, height, or diseases such as **diabetes**, are controlled by a combination of inherited and environmental factors, and stochastic developmental and molecular processes. The easiest traits to dissect genetically are those determined in large part by single genes—so-called **Mendelian** traits. However, almost all of the phenotypic traits of most interest to both anthropologists and physicians are not so simple. Such complex traits are governed by interactions between multiple genes and the environment, and disentangling these interactions will help to relieve the considerable burden of complex diseases on individuals and economies.

Knowledge of our past helps us to predict the numbers and frequencies of genetic variants that influence a given trait and to choose the best strategy for identifying them—how best to define a human population (**Box 1.1**), which populations to choose, and which segments of the genome to concentrate on.

Box 1.1: Caucasians, Caucasoids, European-Americans, Whites? The confusing classification of human social groups

Population geneticists, forensic geneticists, anthropologists, and archaeologists need labels to refer to social groups of human beings. These labels differ between fields, and even within fields. In papers describing DNA diversity in a group of people living in the USA whose ancestors came from Europe, for example, you may find them referred to as:

• A ***US population***: because of population admixture in the last 500 years, the people of the USA are an extremely heterogeneous group, of which those of European descent form only a part.

• ***Caucasians***, or ***Caucasoids***: this is not meant to imply an origin in the Caucasus mountains, but refers to "beautiful people" in a racial classification scheme of the German anatomist Blumenbach (1752–1840). The skull that was claimed to best represent the characteristics of this group came from the Caucasus. The other classifications in this scheme were Mongoloid, Malay, Ethiopian, and American (referring to Native Americans).

• ***European-Americans***: usually in contradistinction to African-, Hispanic-, Japanese-, Native-, and other Americans. There is heterogeneity within this grouping—people who would classify themselves as European-American have ancestors from many different parts of Europe, such as Ireland, Italy, Poland, Russia, and Turkey.

• ***Whites***: this classification is favored by some scientific journals over "Caucasians," which might quite reasonably be reserved for people who really *do* come from the Caucasus. However, it seems odd to use "Whites," when authors of a paper may have no idea what skin color the donors of their DNA samples had.

Often these racial or ethnic labels disguise a great deal of biological heterogeneity, and the identification of DNA donors as members of social groups is not self-evident. Much confusion is possible when a paper has the title: "Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia"[2]. It includes labels based on indigenous continental affiliation, skin color, membership of a group defined on religious-historical grounds, and current small-scale geography. A DNA donor may belong to a large number of categories simultaneously. In addition, when populations are compared in genetic studies, the level of classification in different samples may be unequal. The Hadza of Tanzania, with a population size of only 1000, have in some studies occupied the same analytical status as the South Chinese, whose population size is 600,000 times greater.

In general, the most suitable default method for classification is to use geographical information, rather than national, cultural, or phenotypic labels.