HUMAN EVOLUTIONARY GENETICS

second edition

HUMAN EVOLUTIONARY GENETICS

second edition

Mark Jobling

Edward Hollox

Matthew Hurles

Toomas Kivisild

Chris Tyler-Smith

GS 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

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

Chapter 1	An Introduction to Human Evolutionary Genetics	1
Chapter 2	Organization and Inheritance of the Human Genome	17
Chapter 3	Human Genome Variation	43
Chapter 4	Finding and Assaying Genome Diversity	95
Chapter 5	Processes Shaping Diversity	133
Chapter 6	Making Inferences from Diversity	167
Chapter 7	Humans as Apes	225
Chapter 8	What Genetic Changes Have Made Us Human?	257
Chapter 9	Origins of Modern Humans	283
Chapter 10	The Distribution of Diversity	319
Chapter 11	The Colonization of the Old World and Australia	341
Chapter 12	Agricultural Expansions	363
Chapter 13	Into New-Found Lands	409
Chapter 14	What Happens When Populations Meet	443
Chapter 15	Understanding the Past, Present, and Future of Phenotypic Variation	477
Chapter 16	Evolutionary Insights into Simple Genetic Diseases	517
Chapter 17	Evolution and Complex Diseases	541
Chapter 18	Identity and Identification	571
Appendix		601
Glossary		609
Index		641

DETAILED CONTENTS

CHAPTER 1 AN INTRODUCTION TO HUMAN EVOLUTIONARY GENETICS 1

1.1 WHAT IS HUMAN EVOLUTIONARY GENETICS? 1

1.2 INSIGHTS INTO PHENOTYPES AND DISEASES 2

A shared evolutionary history underpins our understanding of biology 2

Understanding evolutionary history is essential to understanding human biology today 4

Understanding evolutionary history shapes our expectations about the future 5

1.3 COMPLEMENTARY RECORDS OF THE HUMAN PAST 6

Understanding chronology allows comparison of evidence from different scientific approaches 8

It is important to synthesize different records of the past 10

None of the different records represents an unbiased picture of the past 10

1.4 WHAT CAN WE KNOW ABOUT THE PAST? 11

1.5 THE ETHICS OF STUDYING HUMAN POPULATIONS 12

SUMMARY 14

REFERENCES 14

CHAPTER 2 ORGANIZATION AND INHERITANCE OF THE HUMAN GENOME 17

2.1 THE BIG PICTURE: AN OVERVIEW OF THE HUMAN GENOME 17

2.2 STRUCTURE OF DNA 20

2.3 GENES, TRANSCRIPTION, AND TRANSLATION 22

Genes are made up of introns and exons, and include elements to initiate and regulate transcription 22

The genetic code allows nucleotide sequences to be translated into amino acid sequences 24

Gene expression is highly regulated in time and space 26

2.4 NONCODING DNA 26

Some DNA sequences in the genome are repeated in multiple copies 27

2.5 HUMAN CHROMOSOMES AND THE HUMAN KARYOTYPE 28

The human genome is divided into 46 chromosomes 29

Size, centromere position, and staining methods allow chromosomes to be distinguished 31

2.6 MITOSIS, MEIOSIS, AND THE INHERITANCE OF THE GENOME 31

2.7 RECOMBINATION—THE GREAT RESHUFFLER 34

2.8 NONRECOMBINING SEGMENTS OF THE GENOME 36

The male-specific Y chromosome escapes crossing over for most of its length 37

Maternally inherited mtDNA escapes from recombination 37

SUMMARY 40

QUESTIONS 41

REFERENCES 41

CHAPTER 3 HUMAN GENOME VARIATION 43

3.1 GENETIC VARIATION AND THE PHENOTYPE 43

Some DNA sequence variation causes Mendelian genetic disease 44

The relationship between genotype and phenotype is usually complex 46

Mutations are diverse and have different rates and mechanisms 46

3.2 SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS) IN THE NUCLEAR GENOME 47

Base substitutions can occur through base misincorporation during DNA replication 49

Base substitutions can be caused by chemical and physical mutagens 51

Sophisticated DNA repair processes can fix much genome damage 52

The rate of base substitution can be estimated indirectly or directly 53

Because of their low mutation rate, SNPs usually show identity by descent 55

The CpG dinucleotide is a hotspot for mutation 55

Base substitutions and indels can affect the functions of genes 57

Synonymous base substitutions 57

Nonsynonymous base substitutions 58

Indels within genes 59

Base substitutions outside ORFs 60

Whole-genome resequencing provides an unbiased picture of SNP diversity 61

3.3 SEQUENCE VARIATION IN MITOCHONDRIAL DNA 62

mtDNA has a high mutation rate 62

The transmission of mtDNA mutations between generations is complex 64

3.4 VARIATION IN TANDEMLY REPEATED DNA SEQUENCES 65

Microsatellites have short repeat units and repeat arrays, and mutate through replication slippage 66

Microsatellite mutation rates and processes 67

Minisatellites have longer repeat units and arrays, and mutate through recombination mechanisms 69

Minisatellite diversity and mutation 70

Telomeres contain specialized and functionally important repeat arrays 71

Satellites are large, sometimes functionally important, repeat arrays 72

3.5 TRANSPOSABLE ELEMENT INSERTIONS 73

3.6 STRUCTURAL VARIATION IN THE GENOME 75

Some genomic disorders arise from recombination between segmental duplications 76

Copy-number variation is widespread in the human genome 77

Cytogenetic examination of chromosomes can reveal large-scale structural variants 78

3.7 THE EFFECTS OF AGE AND SEX ON MUTATION RATE 78

3.8 THE EFFECTS OF RECOMBINATION ON GENOME VARIATION 81

Genomewide haplotype structure reveals past recombination behavior 84

Recombination behavior can be revealed by direct studies in pedigrees and sperm DNA 87

The process of gene conversion results in nonreciprocal exchange between DNA sequences 88

SUMMARY 90

QUESTIONS 91

REFERENCES 92

CHAPTER 4 FINDING AND ASSAYING GENOME DIVERSITY 95

4.1 FIRST, FIND YOUR DNA 96

4.2 THE POLYMERASE CHAIN REACTION (PCR) 98

4.3 SANGER SEQUENCING, THE HUMAN REFERENCE SEQUENCE, AND SNP DISCOVERY 100

4.4 A QUANTUM LEAP IN VARIATION STUDIES: NEXT-GENERATION SEQUENCING 101

Illumina sequencing is a widely used NGS method 102

Sequencing can be targeted to regions of specific interest or the exome 105

NGS data have to be processed and interpreted 106

Third-generation methods use original, unamplified DNA 107

4.5 SNP TYPING: LOW-, MEDIUM-, AND HIGH-THROUGHPUT METHODS FOR ASSAYING VARIATION 108

PCR-RFLP typing is a simple low-throughput method 108

Primer extension and detection by mass spectrometry is a medium-throughput method 109

High throughput SNP chips simultaneously analyze more than 1 million SNPs 110

Whole-genome SNP chips are based on a tag SNP design 110

4.6 DATABASES OF SEQUENCE VARIATION 112

4.7 DISCOVERING AND ASSAYING VARIATION AT MICROSATELLITES 112

4.8 DISCOVERING AND ASSAYING STRUCTURAL VARIATION ON DIFFERENT SCALES 114

Discovering and assaying variation at minisatellites 114

Discovering and assaying variation at well-defined indels, including *Alu*/LINE polymorphisms 115

Discovering and assaying structural polymorphisms and copy-number variants 115

4.9 PHASING: FROM GENOTYPES TO HAPLOTYPES 119

Haplotypes can be determined by physical separation 120

Haplotypes can be determined by statistical methods 120

Haplotypes can be determined by pedigree analysis 122

4.10 STUDYING GENETIC VARIATION IN ANCIENT SAMPLES 123

DNA is degraded after death 123

Contamination is a major problem 125

Application of next-generation sequencing to aDNA analysis 127

SUMMARY 129

QUESTIONS 130

REFERENCES 130

CHAPTER 5 PROCESSES SHAPING DIVERSITY 133

5.1 BASIC CONCEPTS IN POPULATION GENETICS 133

Why do we need evolutionary models? 133

The Hardy–Weinberg equilibrium is a simple model in population genetics 134

5.2 GENERATING DIVERSITY BY MUTATION AND RECOMBINATION 136

Mutation changes allele frequencies 137

Mutation can be modeled in different ways 137

Meiotic recombination generates new combinations of alleles 139

Linkage disequilibrium is a measure of recombination at the population level 140

Recombination results in either crossing over or gene conversion, and is not uniform across the genome 140

5.3 ELIMINATING DIVERSITY BY GENETIC DRIFT 141

The effective population size is a key concept in population genetics 142

Different parts of the genome have different effective population sizes 143
Genetic drift causes the fixation and elimination of new alleles 143
Variation in census population size and reproductive success influence effective population size 144
Population subdivision can influence effective population size 147
Mate choice can influence effective population size 148
Genetic drift influences the disease heritages of isolated populations 149

5.4 THE EFFECT OF SELECTION ON DIVERSITY 149

Mate choice can affect allele frequencies by sexual selection 153

5.5 MIGRATION 154

There are several models of migration 154
There can be sex-specific differences in migration 155

5.6 INTERPLAY AMONG THE DIFFERENT FORCES OF EVOLUTION 156

There are important equilibria in population genetics 157
Mutation–drift balance 157
Recombination–drift balance 157
Mutation–selection balance 158
Does selection or drift determine the future of an allele? 159

5.7 THE NEUTRAL THEORY OF MOLECULAR EVOLUTION 160

The molecular clock assumes a constant rate of mutation and can allow dating of speciation 160
There are problems with the assumptions of the molecular clock 161

SUMMARY 163

QUESTIONS 164

REFERENCES 164

CHAPTER 6 MAKING INFERENCES FROM DIVERSITY 167

6.1 WHAT DATA CAN WE USE? 167

6.2 SUMMARIZING GENETIC VARIATION 168

Heterozygosity is commonly used to measure genetic diversity 168
Nucleotide diversity can be measured using the population mutation parameter theta (θ) 169
The mismatch distribution can be used to represent genetic diversity 172

6.3 MEASURING GENETIC DISTANCE 173

Genetic distances between populations can be measured using F_{ST} or Nei's *D* statistics 173
Distances between alleles can be calculated using models of mutation 175
Genomewide data allow calculation of genetic distances between individuals 176
Complex population structure can be analyzed statistically 177
Population structure can be analyzed using genomic data 178
Genetic distance and population structure can be represented using multivariate analyses 179

6.4 PHYLOGENETICS 182

Phylogenetic trees have their own distinctive terminology 182
There are several different ways to reconstruct phylogenies 184
Trees can be constructed from matrices of genetic distances 184
Trees can be generated using character-based methods 185
How confident can we be of a particular phylogenetic tree? 188
Networks are methods for displaying multiple equivalent trees 188

6.5 COALESCENT APPROACHES TO RECONSTRUCTING POPULATION HISTORY 190

The genealogy of a DNA sequence can be described mathematically 191
Neutral mutations can be modeled on the gene genealogy using Poisson statistics 192
Coalescent analysis can be a simulation tool for hypothesis testing 193
Coalescent analysis uses ancestral graphs to model selection and recombination 193
Coalescent models of large datasets are approximate 194

6.6 DATING EVOLUTIONARY EVENTS USING GENETIC DATA 194

Dating population splits using F_{ST} and Nei's *D* statistics is possible, but requires a naive view of human evolution 195
Evolutionary models can include the timing of evolutionary events as parameters 195
Evolutionary models and effective population size 196
An allele can be dated using diversity at linked loci 197
Interpreting TMRCA 198
Estimations of mutation rate can be derived from direct measurements in families or indirect comparisons of species 198
An estimate of generation time is required to convert some genetic date estimates into years 198

6.7 HAS SELECTION BEEN ACTING? 200

Differences in gene sequences between species can be used to detect selection 203
Comparing variation *between* species with variation *within* a species can detect selection 207
Selection tests can be based on the analysis of allele frequencies at variant sites 208

Comparing haplotype frequency and haplotype diversity can reveal positive selection 209
Analysis of frequency differences between populations can indicate positive selection 209
Other methods can be used to detect ongoing or very recent positive selection 214
How can we combine information from different statistical tests? 214
Tests for positive selection have severe limitations 215

6.8 ANALYZING GENETIC DATA IN A GEOGRAPHICAL CONTEXT 216

Genetic data can be displayed on maps 217
Genetic boundary analysis identifies the zones of greatest allele frequency change within a genetic landscape 219
Spatial autocorrelation quantifies the relationship of allele frequency with geography 219
Mantel testing is an alternative approach to examining a relationship between genetic distance and other distance measures 220

SUMMARY 220
QUESTIONS 221
REFERENCES 222

CHAPTER 7 HUMANS AS APES 225

Which nonhuman animals are the closest living relatives of humans? 225
Are humans typical apes? 225

7.1 EVIDENCE FROM MORPHOLOGY 226

Primates are an Order of mammals 226
Hominoids share a number of phenotypic features with other anthropoids 228
Ancestral relationships of hominoids are difficult to resolve on morphological evidence 230

7.2 EVIDENCE FROM CHROMOSOMES 232

Human and great ape karyotypes look similar, but not identical 232
Molecular cytogenetic analyses support the picture from karyotype comparisons 233

7.3 EVIDENCE FROM MOLECULES 236

Molecular data support a recent date of the ape–human divergence 237
Genetic data have resolved the gorilla–chimpanzee–human trichotomy 237
Sequence divergence is different among great apes across genetic loci 239
Great apes differ by gains and losses of genetic material 241
The DNA sequence divergence rates differ in hominoid lineages 241

7.4 GENETIC DIVERSITY AMONG THE GREAT APES 242

How many genera, species, and subspecies are there? 247
Intraspecific diversity in great apes is greater than in humans 247
Signatures of lineage-specific selection can be detected in ape genomes 250

SUMMARY 254
QUESTIONS 254
REFERENCES 254

CHAPTER 8 WHAT GENETIC CHANGES HAVE MADE US HUMAN? 257

8.1 MORPHOLOGICAL AND BEHAVIORAL CHANGES EN ROUTE TO *HOMO SAPIENS* 258

Some human traits evolved early in hominin history 260
The human mind is unique 263
Only a few phenotypes are unique to modern humans 265

8.2 GENETIC UNIQUENESS OF HUMANS AND HOMININS 265

The sequence and structural differences between humans and other great apes can be cataloged 265
Humans have gained and lost a few genes compared with other great apes 266
Humans differ in the sequence of genes compared with other great apes 269
Humans differ from other apes in the expression levels of genes 270
Genome sequencing has revealed a small number of fixed genetic differences between humans and both Neanderthals and Denisovans 272

8.3 GENETIC BASIS OF PHENOTYPIC DIFFERENCES BETWEEN APES AND HUMANS 273

Mutations causing neoteny have contributed to the evolution of the human brain 273
The genetic basis for laterality and language remains unclear 275
What next? 278

SUMMARY 278
QUESTIONS 279
REFERENCES 279

CHAPTER 9 ORIGINS OF MODERN HUMANS 283

9.1 EVIDENCE FROM FOSSILS AND MORPHOLOGY 284

Some fossils that may represent early hominins from 4–7 MYA are known from Africa 285
Fossils of australopithecines and their contemporaries are known from Africa 287
The genus *Homo* arose in Africa 290
The earliest anatomically modern human fossils are found in Africa 294
The morphology of current populations suggests an origin in Africa 295

9.2 EVIDENCE FROM ARCHAEOLOGY AND LINGUISTICS 295

Paleolithic archaeology has been studied extensively 298

Evidence from linguistics suggests an origin of language in Africa 299

9.3 HYPOTHESES TO EXPLAIN THE ORIGIN OF MODERN HUMANS 300

9.4 EVIDENCE FROM THE GENETICS OF PRESENT-DAY POPULATIONS 301

Genetic diversity is highest in Africa 301

Genetic phylogenies mostly root in Africa 304

Mitochondrial DNA phylogeny 304

Y-chromosomal phylogeny 305

Other phylogenies 305

Insights can be obtained from demographic models 306

9.5 EVIDENCE FROM ANCIENT DNA 307

Ancient mtDNA sequences of Neanderthals and Denisovans are distinct from modern human variation 308

A Neanderthal draft genome sequence has been generated 309

A Denisovan genome sequence has been generated 310

SUMMARY 313

QUESTIONS 315

REFERENCES 315

CHAPTER 10 THE DISTRIBUTION OF DIVERSITY 319

10.1 STUDYING HUMAN DIVERSITY 319

The history and ethics of studying diversity are complex 319

Linnaeus' classification of human diversity 320

Galton's "Comparative worth of different races" 320

Modern attitudes to studying diversity 320

Who should be studied? 323

A few large-scale studies of human genetic variation have made major contributions to human evolutionary genetics 323

What is a population? 326

How many people should be analyzed? 327

10.2 APPORTIONMENT OF HUMAN DIVERSITY 328

The apportionment of diversity shows that most variation is found within populations 328

The apportionment of diversity can differ between segments of the genome 329

Patterns of diversity generally change gradually from place to place 330

The origin of an individual can be determined surprisingly precisely from their genotype 331

The distribution of rare variants differs from that of common variants 332

10.3 THE INFLUENCE OF SELECTION ON THE APPORTIONMENT OF DIVERSITY 333

The distribution of levels of differentiation has been studied empirically 334

Low differentiation can result from balancing selection 334

High differentiation can result from directional selection 335

Positive selection at EDAR 336

SUMMARY 338

QUESTIONS 339

REFERENCES 339

CHAPTER 11 THE COLONIZATION OF THE OLD WORLD AND AUSTRALIA 341

11.1 A COLDER AND MORE VARIABLE ENVIRONMENT 15–100 KYA 341

11.2 FOSSIL AND ARCHAEOLOGICAL EVIDENCE FOR TWO EXPANSIONS OF ANATOMICALLY MODERN HUMANS OUT OF AFRICA IN THE LAST ~130 KY 344

Anatomically modern, behaviorally pre-modern humans expanded transiently into the Middle East ~90–120 KYA 345

Modern human behavior first appeared in Africa after 100 KYA 346

Fully modern humans expanded into the Old World and Australia ~50–70 KYA 347

Modern human fossils in Asia, Australia, and Europe 347

Initial colonization of Australia 349

Upper Paleolithic transition in Europe and Asia 352

11.3 A SINGLE MAJOR MIGRATION OUT OF AFRICA 50–70 KYA 353

Populations outside Africa carry a shared subset of African genetic diversity with minor Neanderthal admixture 353

mtDNA and Y-chromosomal studies show the descent of all non-African lineages from a single ancestor for each who lived 55–75 KYA 355

11.4 EARLY POPULATION DIVERGENCE BETWEEN AUSTRALIANS AND EURASIANS 357

SUMMARY 360

QUESTIONS 361

REFERENCES 361

CHAPTER 12 AGRICULTURAL EXPANSIONS 363

12.1 DEFINING AGRICULTURE 363

12.2 THE WHERE, WHEN, AND WHY OF AGRICULTURE 365

Where and when did agriculture develop? 365

Why did agriculture develop? 366

Which domesticates were chosen? 368

12.3 OUTCOMES OF AGRICULTURE 369

Agriculture had major impacts on demography and disease 369

Rapid demographic growth 369

Malnutrition and infectious disease 369
Agriculture led to major societal changes 371

12.4 THE FARMING–LANGUAGE CO-DISPERSAL HYPOTHESIS 372

Some language families have spread widely and rapidly 372
Linguistic dating and construction of proto-languages have been used to test the hypothesis 373
What are the genetic implications of language spreads? 373

12.5 OUT OF THE NEAR EAST INTO EUROPE 374

Nongenetic evidence provides dates for the European Neolithic 374
Different models of expansion give different expectations for genetic patterns 377
Models are oversimplifications of reality 378
Principal component analysis of classical genetic polymorphisms was influential 379
Interpreting synthetic maps 379
mtDNA evidence has been controversial, but ancient DNA data are transforming the field 380
Data from ancient mtDNA 382
Y-chromosomal data show strong clines in Europe 384
New developments for the Y chromosome 384
Biparentally inherited nuclear DNA has not yet contributed much, but important ancient DNA data are now emerging 386
Ancient DNA data 387
What developments will shape debate in the future? 388

12.6 OUT OF TROPICAL WEST AFRICA INTO SUB-EQUATORIAL AFRICA 388

There is broad agreement on the background to African agricultural expansion 388
Rapid spread of farming economies 389
Bantu languages spread far and rapidly 390
Genetic evidence is broadly consistent, though ancient DNA data are lacking 392
Genomewide evidence 392
Evidence from mtDNA and the Y chromosome 393

12.7 GENETIC ANALYSIS OF DOMESTICATED ANIMALS AND PLANTS 394

Selective regimes had a massive impact on phenotypes and genetic diversity 395
Key domestication changes in crops 396
Effects on crop genetic diversity 398
Phenotypic and genetic change in animals 399
How have the origins of domesticated plants been identified? 400
How have the origins of domesticated animals been identified? 401
Cattle domestication 403

SUMMARY 404
QUESTIONS 405
REFERENCES 405

CHAPTER 13 INTO NEW-FOUND LANDS 409

13.1 SETTLEMENT OF THE NEW TERRITORIES 409

Sea levels have changed since the out-of-Africa migration 409
What drives new settlement of uninhabited lands? 411

13.2 PEOPLING OF THE AMERICAS 412

The changing environment has provided several opportunities for the peopling of the New World 413
Fossil and archaeological evidence provide a range of dates for the settlement of the New World 415
Fossils 415
Archaeological remains 416
Clovis and the Paleoindians 416
Pre-Clovis sites 416
Unresolved issues 417
Did the first settlers go extinct? 418
A three-migration hypothesis has been suggested on linguistic grounds 419
Genetic evidence has been used to test the single- and the three-wave migration scenarios 419
Mitochondrial DNA evidence 420
Interpretation of the mtDNA data 422
Evidence from the Y chromosome 422
Evidence from the autosomes 424
Conclusions from the genetic data 425

13.3 PEOPLING OF THE PACIFIC 425

Fossil and archaeological evidence suggest that Remote Oceania was settled more recently than Near Oceania 427
Two groups of languages are spoken in Oceania 428
Several models have been proposed to explain the spread of Austronesian speakers 430
Austronesian dispersal models have been tested with genetic evidence 431
Classical polymorphisms 431
Globin gene mutations 432
Mitochondrial DNA 433
The Y chromosome 436
Autosomal evidence 437
Evidence from other species has been used to test the Austronesian dispersal models 438

SUMMARY 440
QUESTIONS 441
REFERENCES 441

CHAPTER 14 WHAT HAPPENS WHEN POPULATIONS MEET 443

14.1 WHAT IS GENETIC ADMIXTURE? 443

Admixture has distinct effects on genetic diversity 445

14.2 THE IMPACT OF ADMIXTURE 447

Different sources of evidence can inform us about admixture 447

Consequences of admixture for language 447
Archaeological evidence for admixture 448
The biological impact of admixture 449

14.3 DETECTING ADMIXTURE 450

Methods based on allele frequency can be used to detect admixture 450
Admixture proportions vary among individuals and populations 453
Calculating individual admixture levels using multiple loci 453
Calculating individual admixture levels using genomewide data 454
Calculating admixture levels from estimated ancestry components 456
Problems of measuring admixture 457
Natural selection can affect the admixture proportions of individual genes 458

14.4 LOCAL ADMIXTURE AND LINKAGE DISEQUILIBRIUM 460

How does admixture generate linkage disequilibrium? 461
Admixture mapping 462
Admixture dating 463

14.5 SEX-BIASED ADMIXTURE 464

What is sex-biased admixture? 464
Detecting sex-biased admixture 465
Sex-biased admixture resulting from directional mating 465
The effect of admixture on our genealogical ancestry 467

14.6 TRANSNATIONAL ISOLATES 467

Roma and Jews are examples of widely spread transnational isolates 468
European Roma 468
The Jews 469

SUMMARY 471
QUESTIONS 472
REFERENCES 473

CHAPTER 15 UNDERSTANDING THE PAST, PRESENT, AND FUTURE OF PHENOTYPIC VARIATION 477

15.1 NORMAL AND PATHOGENIC VARIATION IN AN EVOLUTIONARY CONTEXT 477

15.2 KNOWN VARIATION IN HUMAN PHENOTYPES 478

What is known about human phenotypic variation? 478
Morphology and temperature adaptation 479
Facial features 479
Tooth morphology and cranial proportions 480
Behavioral differences 481
How do we uncover genotypes underlying phenotypes? 483
What have we discovered about genotypes underlying phenotypes? 485

15.3 SKIN PIGMENTATION AS AN ADAPTATION TO ULTRAVIOLET LIGHT 485

Melanin is the most important pigment influencing skin color 486
Variable ultraviolet light exposure is an adaptive explanation for skin color variation 486
Several genes that affect human pigmentation are known 489
Genetic variation in human pigmentation genes is consistent with natural selection 492
Does sexual selection have a role in human phenotypic variation? 493

15.4 LIFE AT HIGH ALTITUDE AND ADAPTATION TO HYPOXIA 495

Natural selection has influenced the overproduction of red blood cells 495
High-altitude populations differ in their adaptation to altitude 496

15.5 VARIATION IN THE SENSE OF TASTE 496

Variation in tasting phenylthiocarbamide is mostly due to alleles of the *TAS2R38* gene 498
There is extensive diversity of bitter taste receptors in humans 499
Sweet, umami, and sour tastes may show genetic polymorphism 499

15.6 ADAPTING TO A CHANGING DIET BY DIGESTING MILK AND STARCH 500

There are several adaptive hypotheses to explain lactase persistence 501
Lactase persistence is caused by SNPs within an enhancer of the lactase gene 502
Increased copy number of the amylase gene reflects an adaptation to a high-starch diet 504

15.7 THE FUTURE OF HUMAN EVOLUTION 506

Have we stopped evolving? 506
Natural selection acts on modern humans 506
Can we predict the role of natural selection in the future? 507
Climate change 507
Dietary change 507
Infectious disease 507
What will be the effects of future demographic changes? 508
Increasing population size 509
Increased mobility 510
Differential fertility 510
Differential generation time 511
Will the mutation rate change? 512

SUMMARY 512
QUESTIONS 513
REFERENCES 513

CHAPTER 16 EVOLUTIONARY INSIGHTS INTO SIMPLE GENETIC DISEASES 517

16.1 GENETIC DISEASE AND MUTATION–SELECTION BALANCE 520

Variation in the strength of purifying selection can affect incidence of genetic disease 520

Variation in the deleterious mutation rate can affect incidence of genetic disease 522

16.2 GENETIC DRIFT, FOUNDER EFFECTS, AND CONSANGUINITY 523

Jewish populations have a particular disease heritage 524

Finns have a disease heritage very distinct from other Europeans 525

Consanguinity can lead to increased rates of genetic disease 526

16.3 EVOLUTIONARY CAUSES OF GENOMIC DISORDERS 526

Segmental duplications allow genomic rearrangements with disease consequences 527

Duplications accumulated in ancestral primates 529

16.4 GENETIC DISEASES AND SELECTION BY MALARIA 529

Sickle-cell anemia is frequent in certain populations due to balancing selection 531

α-Thalassemias are frequent in certain populations due to balancing selection 534

Glucose-6-phosphate dehydrogenase deficiency alleles are maintained at high frequency in malaria-endemic populations 535

What can these examples tell us about natural selection? 537

SUMMARY 538

QUESTIONS 538

REFERENCES 539

CHAPTER 17 EVOLUTION AND COMPLEX DISEASES 541

17.1 DEFINING COMPLEX DISEASE 541

The genetic contribution to variation in disease risk varies between diseases 544

Infectious diseases are complex diseases 544

17.2 THE GLOBAL DISTRIBUTION OF COMPLEX DISEASES 546

Is diabetes a consequence of a post-agricultural change in diet? 546

The drifty gene hypothesis 547

Evidence from genomewide studies 548

The thrifty phenotype hypothesis 549

17.3 IDENTIFYING ALLELES INVOLVED IN COMPLEX DISEASE 549

Genetic association studies are more powerful than linkage studies for detecting small genetic effects 549

Candidate gene association studies have not generally been successful in identifying susceptibility alleles for complex disease 552

Genomewide association studies can reliably identify susceptibility alleles to complex disease 552

GWAS data have been used for evolutionary genetic analysis 556

17.4 WHAT COMPLEX DISEASE ALLELES DO WE EXPECT TO FIND IN THE POPULATION? 557

Negative selection acts on disease susceptibility alleles 557

Positive selection acts on disease resistance alleles 560

Severe sepsis and CASP12 560

Malaria and the Duffy antigen 560

HIV-1 and CCR5Δ32 562

Unexpectedly, some disease susceptibility alleles with large effects are observed at high frequency 562

Susceptibility to kidney disease, APOL1, and resistance to sleeping sickness 562

Implications for other GWAS results 563

17.5 GENETIC INFLUENCE ON VARIABLE RESPONSE TO DRUGS 563

Population differences in drug-response genes exist, but are not well understood 564

SUMMARY 567

QUESTIONS 568

REFERENCES 569

CHAPTER 18 IDENTITY AND IDENTIFICATION 571

18.1 INDIVIDUAL IDENTIFICATION 572

The first DNA fingerprinting and profiling methods relied on minisatellites 573

PCR-based microsatellite profiling superseded minisatellite analysis 574

How do we interpret matching DNA profiles? 574

Complications from related individuals, and DNA mixtures 576

Large forensic identification databases are powerful tools in crime-fighting 577

Controversial aspects of identification databases 577

The Y chromosome and mtDNA are useful in specialized cases 578

Y chromosomes in individual identification 579

mtDNA in individual identification 580

18.2 WHAT DNA CAN TELL US ABOUT JOHN OR JANE DOE 580

DNA-based sex testing is widely used and generally reliable 580

Sex reversal 581

Deletions of the AMELY locus in normal males 582

Some other phenotypic characteristics are predictable from DNA 582

Reliability of predicting population of origin depends on what DNA variants are analyzed 583
Prediction from forensic microsatellite multiplexes 583
Prediction from other systems 584
The problem of admixed populations 584

18.3 DEDUCING FAMILY AND GENEALOGICAL RELATIONSHIPS 585

The probability of paternity can be estimated confidently 586
Other aspects of kinship analysis 588
The Y chromosome and mtDNA are useful in genealogical studies 588
The Thomas Jefferson paternity case 588
DNA-based identification of the Romanovs 590
Y-chromosomal DNA has been used to trace modern diasporas 591
Y-chromosomal haplotypes tend to correlate with patrilineal surnames 592

18.4 THE PERSONAL GENOMICS REVOLUTION 593

The first personal genetic analysis involved the Y chromosome and mtDNA 593
Personal genomewide SNP analysis is used for ancestry and health testing 593
Personal genome sequencing provides the ultimate resolution 593
Personal genomics offers both promise and problems 596

SUMMARY 597

QUESTIONS 597

REFERENCES 598

APPENDIX 601

HAPLOGROUP NOMENCLATURE 601

THE MITOCHONDRIAL GENOME 602

What are its origins? 602
What genes are encoded within the mitochondrial genome? 602
What diseases are caused by mutations within mtDNA? 602
How has the study of mtDNA diversity developed? 602
How is information from the mtDNA variants in an individual combined? 603
Why are all the deep-rooting clades called L? 603
Why is mtDNA so useful for exploring the human past? 603
What about possible selection pressures? 605

THE Y CHROMOSOME 605

How has it evolved? 605
What does the chromosome contain? 605
How similar are Y chromosomes within and between species? 606
What molecular polymorphisms are found on the Y chromosome? 606
How should the polymorphic information from different variants be combined? 606
What are the applications of studying Y-chromosomal diversity? 608
Is there any evidence of selection on the Y chromosome? 608

REFERENCES 608

GLOSSARY 609

INDEX 641