Foundations of Human Biology • Matt Cartmill and Kaye Brown, Series Editors

HUMAN POPULATION GENETICS

John H. Relethford

WILEY-BLACKWELL

HUMAN POPULATION GENETICS

Foundations of Human Biology

Series Editors:
Matt Cartmill
Kaye Brown
Boston University

The Growth of Humanity by Barry Bogin

Fundamentals of Forensic Anthropology by Linda L. Klepinger

The Human Lineage by Matt Cartmill and Fred H. Smith

HUMAN POPULATION GENETICS

JOHN H. RELETHFORD
Department of Anthropology
State University of New York College at Oneonta

WILEY-BLACKWELL

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical, and Medical business with Blackwell Publishing.

Library of Congress Cataloging-in-Publication Data:

Relethford, John.
Human population genetics / John H. Relethford.
p. cm.
Includes index.
Summary: ''*Human Population Genetics* will provide an introduction to mathematical population genetics, along with relevant examples from human (and some non-human primate) populations, and will also present concepts and methods of population genetics that are specific to the study of human populations. The purpose of this book is to provide a basic background text for advanced undergraduate and graduate students interesting in the mechanisms of human microevolution''—Provided by publisher.
ISBN 978-0-470-46467-0 (pbk.)
1. Human population genetics. I. Title.
GN289.R45 2012
599.93'5—dc23

2011028962

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

FOREWORD

If, like us, you find yourself hard-pressed to follow the fast-paced scrimmages of anthropological genetics from the sidelines, this is the book you have been waiting for. John Relethford, one of the world's leading contributors to these debates, has written it to engage all of us in this important and rapidly evolving area of scientific inquiry. In *Human Population Genetics*, he leads us through classic studies and current debates in an easy, clear, informal style that draws us in and involves us in the action and arguments. Relethford's passion for understanding the genetics of human populations, and his low-stress approach to what can be a difficult and esoteric topic, kindle a like passion in the reader and make this book that rare thing among textbooks—a source of excitement and inspiration.

Population genetics and statistical theory were born as conjoined twins in the monumental work of R. A. Fisher in the 1920s, which transformed evolutionary biology into a full-fledged science capable of making and testing predictions with numbers in them. But many people who are eager to learn about human biology and evolution are turned off by the statistical foundations of evolutionary theory. Almost everyone who teaches the fundamentals of our science has learned to dread the dazed expressions that come over students' faces the moment the Hardy–Weinberg equation hits the screen. Relethford shows us, and them, how to get around this stumbling block. Drawing the reader effortlessly in through plain and simple examples beautifully chosen to clarify the mathematics of probability, Relethford recruits his mastery of the subject and his skill as a teacher and writer to present the math in a user-friendly way that displaces the hard work of deriving formulas into adjacent appendices. His readers first master the essentials and later reward themselves by seeing the mathematics underlying the simple models they have just grasped. This process of orderly presentation leaves readers self-confident and ready to take on ever more complex material.

Throughout this book, Relethford systematically preaches and teaches a scientific approach to knowledge (''*Much of science consists of developing a simple model, testing its fit in the real world, and then explaining why and how it fits and does not fit*'') in a way that always solicits involvement by the reader (''*To see this, let us try an example*''). In every topic he presents, he returns to the readers' point of view (''*What effect do you think selection has had on the allele frequencies*?'') and includes them in the developing narrative. His readers will learn the concepts that are crucial to all fields of population biology by studying examples of special relevance to biological anthropology—how familiarity with genetic evidence can

inform us of our history (see the rich discussion on tracking the appearance of the *CCR5-Δ32* allele and subsequent resistance to the AIDS virus), how adaptation has taken many different paths in human history (see the discussion on different high-altitude adaptations in Tibetan and Andean people), and how cultural behavior impacts genetic processes (see the discussion on agriculture and hemoglobin S). ''Instead of cultural evolution negating genetic evolution,'' he writes, ''we are finding evidence of how cultural change has accelerated genetic evolution.''

That sentence, and the evidence behind it, would by itself make *Human Population Genetics* worth having on your bookshelf. Every chapter of the book sparkles with conclusions that are just as simple, straightforward, and far-reaching. All of its readers can rely on John Relethford to lead them into some of the most important and exciting scientific conversations of our day. If you are a student of biological anthropology at any level, or a scientist or educator who teaches these subjects, you will find his new book an invaluable source of novel insights and fresh illumination of key ideas. We are proud and delighted to see *Human Population Genetics* added to the Wiley–Blackwell series of textbooks on the foundations of human biology.

Kaye Brown
Matt Cartmill

PREFACE

WHAT ARE WE DOING HERE?

This book is about the intersection of mathematics, biology, and anthropology. As such, it has two basic goals. First, the book provides an *introduction* to the study of population genetics, which provides the mathematical basis of evolutionary theory by describing changes in the frequency of genetic variants from one generation to the next. Second, this introduction has been designed for specific application to *human* populations. Although population genetics is a field that applies to all organisms, the focus throughout this book, particularly in case studies, is on human populations. As an anthropologist, my interest is by definition primarily on *human* populations and genetic diversity. Not that this book has no utility outside of human populations—far from it. I have designed this book to provide a simple introduction to population genetics with minimal mathematics that can be used by advanced undergraduate and graduate students in a variety of fields, including anthropology, biology, and ecology. If you are using this book in one of those other disciplines, rest assured that the same basic principles presented here are applicable to organisms, and your instructor will likely provide other, nonhuman, case studies for clarification. You need not have a detailed background in genetics, although this book is intended for students that have had some initial grounding in genetics, such as one would obtain from an introductory course in biological anthropology or biology.

FORMAT AND ORGANIZATION OF THE BOOK

A quick look through the pages of this book will reveal a number of formulas. This may seem intimidating, but it is not. Although some elementary mathematics is needed to understand population genetics, we do not have to use very advanced math to learn the basics. Throughout this book, we will use only simple algebra of the type that you likely learned in middle or senior high school and some basic concepts of probability, which are developed in the text as we proceed. I also use additional ways, beyond equations, to present the material. Although it is a wonderful experience to glance at a mathematical formula and gain immediate insight into what that formula says about reality, it is (at least for me) a rare

experience. I usually have to look at a graphic representation of the formula or utilize an analogy to understand the underlying ideas. Thus, this text uses a lot of graphs and analogies to make the basic points and help you relate the evolutionary process to mathematical ideas.

As with any field, population genetics has its own set of terms. Anything specific to genetics or population genetics is defined in the text, with an additional glossary at the end of the book collecting all such terms. All glossary terms are marked in **boldface** in the text the first time they appear. In-text citation is used in this text, where specific citations are references by author(s) name(s) and year, such as ''Relethford (2004).''

ACKNOWLEDGMENTS

I owe much thanks to Matt Cartmill and Kaye Brown, series editors of the Wiley-Blackwell *Foundation of Human Biology* series, for inviting me to write this book, and for their careful analysis and discussion of the book's goals and structure. I am also very grateful for the guidance and advice of my editor, Karen Chambers. She was a delight to work with on this project. Thanks also to Anna Ehler, Editorial Assistant, and Rosalyn Farkas, Production Editor, for all of their help and attention to my constant questions.

I was first introduced to the study of population genetics in 1975 when I met my graduate school advisor, Frances Lees. I owe Frank a lot for his guidance and friendship over the years in addition to his patience at teaching me population genetics. He got me started both in my profession and in this particular field. I am also very grateful to his academic advisor, Michael Crawford, for helping me learn even more about population genetics over the course of several decades of friendship and collaboration on research projects.

I have worked with other colleagues on research in human population genetics. Two of these colleagues stand out in particular—John Blangero and Henry Harpending. My work with them has been a high point of my career.

Looking back, I can identify many friends and colleagues over the years with whom I have shared discussions at some level or another on population genetics. Some of these have been coauthors, and others have been colleagues with similar interests who have shared one or many conversations or emails. They all have contributed to my understanding of human population genetics. Needless to say, my errors are mine and mine alone. This is the list (and my most sincere apologies if I have missed anyone): Guido Barbujani, Deborah Bolnick, the late Ellen Brennan, Ranajit Chakraborty, Ric Devor, Ravi Duggarali, Elise Eller, Alan Fix, Jon Friedlaender, Rosalind Harding, Mike Hammer, John Hawks, Jeff Heilveil, Keith Hunley, Cashell Jaquish, Lynn Jorde, Lyle Konigsberg, Tibor Koertvelyessy, Ken Korey, the late Gabe Lasker, Paul Leslie, Jeff Long, Lorena Madrigal, Andrea Manica, Yoshiro Matsuo, Jim Mielke, Andy Merriwether, John Mitchell, Kari North, Carolyn Olsen, Esteban Parra, Alan Rogers, Charles Roseman, Dennis O'Rourke, Lisa Sattenspiel, Michael Schillaci, Tad Schurr, Steve Sherry, Peter Smouse, Bob Sokal, Dawnie Steadman, Anne Stone, Mark Stoneking, Alan Swedlund, Alan Templeton, Forrest Tierson, John VandeBerg, Noreen von

Cramon-Taubadel, Tim Weaver, Ken Weiss, Dick Wilkinson, Sarah Williams-Blangero, Milford Wolpoff and Jim Wood. Special thanks to Alan Bittles for providing me with references on inbreeding. I also acknowledge my debt to three individuals whom I have never met, but have spent many hours studying their insightful writings: Luca Cavalli-Sforza, Newton Morton, and the late Sewall Wright.

Last, but certainly not least, I dedicate this book to the five people who mean the most to me in the world—my wife, Hollie Jaffe; my sons, David, Ben, and Zane; and my mother-in-law, Terry Adler. Thanks to all for putting up with me and loving me.

John H. Relethford
State University of New York

CHAPTER 1

GENETIC, MATHEMATICAL, AND ANTHROPOLOGICAL BACKGROUND

My interest in human population genetics started with my difficulty in picking a major in college.

As is often the case, my interests as an undergraduate student were varied, including fields as different as sociology, biology, geography, history, and mathematics. Each of these fields appealed to me in some ways initially, but none sufficiently to take the 10 or more courses to complete an academic major. As I shifted almost daily in my search for a major, I stumbled across anthropology, a discipline that is characterized by academic breadth across the liberal arts. In the United States, anthropology departments are most often constructed around the four-field approach championed by the famous early twentieth-century anthropologist, Franz Boas. Here, anthropology is divided into four subfields: (1) *cultural anthropology*, which examines behaviors in current and recent human populations; (2) *archaeology*, which reconstructs cultural behavior in prehistoric and historic human societies; (3) *linguistics*, the study of language, a uniquely human form of communicating culture; and (4) *biological anthropology* (also known as *physical anthropology*), which focuses on the biological evolution and variation of the human species.

With its focus on both cultural and biological aspects of humanity, and its concern with natural science, social science, and the humanities, anthropology proved to be the perfect liberal arts major for someone like me, who had a difficult time picking any single major. Over time, however, I found myself gravitating more toward the subfield of biological anthropology as I became fascinated by the ways in which humanity had evolved. As I entered graduate school, I wound up concentrating more and more on the nature of human biological variation, and questions about our species' biological diversity. How are human populations similar to and different from each other biologically? How do these differences relate to the process of evolution, and how do these processes relate to human history, culture, and the environment? In one form or another, these

Human Population Genetics, First Edition. John H. Relethford.

questions have been at the root of many of the research topics I have focused on during my career, ranging from the effect of historical invasions on genetic diversity in Ireland, to changing patterns of marriage and migration in colonial Massachusetts, to the effect of history and geography on cranial shape across the world.

Underlying all of these questions is the subject of this book, human population genetics, which is a field that has the same breadth of topics that guided my search for a college major. Although this book focuses on *human* population genetics, it is important to realize that population genetics is a subject that concerns *all* organisms. Much of this book consists in explaining basic principles of population genetics, applicable to many species, with further illustration describing case studies from human populations. If you are reading this book in a course on general population genetics, as is often taught in biology departments, for example, you are likely to encounter further case studies on a variety of other species.

I. THE SCOPE OF POPULATION GENETICS

Before getting too far into the application of population genetics to the human species, it is useful to answer the basic question ''What is population genetics?'' This question can be answered by considering the nature of the broader field of genetics, the study of heredity in organisms. Genetics can be studied at various levels. The study of molecular genetics deals with the biochemical nature of heredity, specifically DNA and RNA. At this level, geneticists focus on the biochemical nature of heredity, including the structure and function of genes and other DNA sequences.

The study of Mendelian genetics, named after the Austrian monk, Gregor Mendel (1822–1884), is concerned with the process and pattern of genetic inheritance from parents to offspring. Mendel's work gave us a basic understanding of how inheritance works, and how discrete units of inheritance combine to produce genotypes and phenotypes. Whereas the focus of molecular genetics is on the transmission of information from cell to cell, Mendelian genetics focuses on the transmission of genetic information from one individual (a parent) to another (the offspring). Mendelian genetics is in essence a statistical subject, dealing with the probability of different genotypes and phenotypes in offspring. A classic example concerns two parents, each of which carries one copy of a recessive gene. The principles of probability show that the chance of any given offspring having *two* copies of that gene, one from each parent, is $\frac{1}{4}$. These principles will be reviewed later, but for now, you should just consider that the transmission of genetic information is subject to the laws of probability.

Population genetics takes this concern with the probability of transmitting genetic information from one generation to the next and extends it to the next level, an entire population (or set of populations, or even an entire species). In population genetics, we are concerned with the genetic composition of the entire population, and how this composition can change over time. For example, consider the classic example of the peppered moth in England. This species of moth comes in two forms, a dark-colored form and a light-colored form. Centuries ago, most moths were light-colored, and only about 1% were dark-colored. Dark-colored

moths were rare because they would be more clearly visible against the light color of the tree trunks, making it easier for birds to see them and eat them. Over time, the environment changed, and the frequency of dark-colored moths increased as the frequency of the light-colored moths decreased. Because the color of the moths reflects genetic differences, this observed change is an example of the genetic composition of a species changing over time. Population genetics deals with explaining such changes. In this case, the initial origin of a different form is due to mutation, and the change in moth color over time reflects natural selection, because the environment had shifted following the Industrial Revolution, leading to darker tree trunks, thus creating a situation where dark-colored moths were less likely to be eaten by birds.

When the genetic makeup of a population changes over time, even in a single generation, we have a case of evolution. Population genetics is the branch of genetics that deals with evolutionary change in populations of organisms, and provides the mathematical basis of evolutionary theory. Note that I am using the word *theory* here in the context of the natural sciences, where a theory is a set of hypotheses that have been tested and have withstood the test of time, as compared with the popular use of the word theory as a simple hypothesis. When we speak of evolutionary theory, we are *not* stating that evolution may or may not exist, but instead are referring to a set of principles that explain the *facts* of evolution (in other words, beware of the statement that ''evolution is a theory and not a fact,'' because it is actually *both* a fact and a theory).

Evolution can be viewed over different scales of time and units of analysis. Population genetics deals with changes within a species over relatively short intervals of time, typically on the order of a small number of generations. This type of evolutionary change is also known as **microevolution**, and is contrasted with **macroevolution**, which focuses on the evolution of species and higher levels (genera, families, etc.), and typically deals with geological timescales, ranging from thousands to millions of years. Although macroevolution and microevolution are related in a theoretical sense, there is continued debate over the extent to which long-term macroevolutionary events are a straightforward extrapolation of microevolutionary trends (Simons 2002). The focus of this book is primarily on the theory of microevolution.

Population genetics is concerned with changes in genetic variation over time, that is, genetic differences and similarities. There are two ways of looking at genetic variation: variation *within* populations and variation *between* populations. The former refers to differences and similarities of individuals within a population; the latter refers to average differences between two or more populations. Later chapters will introduce quantitative measures of within-group and between-group variation based on genetic traits, but for the moment, I will use a simple analogy looking at adult human height. Picture yourself in a large classroom filled with students, and imagine that we measured everyone's height. We would use these measurements to compute how much variation existed *within* the classroom. If, for example, everyone in the class were of exactly the same height, there would be *no* variation. If, however, there were differences in height, with everyone being between 5 ft 8 in tall and 5 ft 10 in. tall, then variation would exist because not

everyone would be the same. If everyone were between 5 and 6 ft tall, there would be even more variation.

On the other hand, suppose that we want to compare the height in your classroom with the height in the next classroom. An example would be if the average height in your classroom were 5 ft 9 in. and the average height in the other classroom were 5 ft 8 in. The difference in average height would be 1 in. This difference would be an example of variation *between* groups. If the average height of the two classes were the same, then there would be no variation between groups. In evolutionary terms, we are interested in changes in genetic variation that take place both within and between populations.

By studying genetic change over time and its effects on genetic variation within and between populations, we are able to apply the theory of population genetics to address a wide variety of questions about human variation and evolution. A small sample of such questions (which will be addressed in later chapters) includes

- How much inbreeding occurs in human populations, and what is the effect of this inbreeding?
- What does genetic variation tell us about our species' history?
- Can genetics to be used to trace ancient human migrations?
- Where did the first Americans come from?
- Why do some human populations have high frequencies of the harmful sickle cell allele?
- Are certain genes resistant to acquired immunodeficiency syndrome (AIDS)?
- Why do some small populations differ genetically from their neighbors to such an extent?
- What impact does geography have on our choice of mates?

Even this short list shows that population genetics has relevance to many questions about human biological variation and evolution. In addition, the general principles of population genetics are used to address the same concerns—variation and evolution—in all organisms. In short, population genetics is a key to understanding life. Although this book focuses on human populations (because of my interests and training), never forget that many of the general principles of population genetics apply across the span of life itself.

As noted earlier, the study of human population genetics examines the application of mathematical principles and models to the transmission of genetic information from one generation to the next in human populations. Population genetics can be regarded here as a field that combines genetics, mathematics (especially probability), and anthropology. The remainder of this introductory chapter provides a brief review of some basic principles of genetics and probability, and concludes with a broader consideration of how population genetics applies in an anthropological context.

II. GENETICS BACKGROUND

Considering the nature of this book and its intended audience, one might assume that you are a student in a course on population genetics or a related field. Typically, such students have had some background in some basic concepts of genetics, particularly Mendelian genetics, from high school as well as in an introductory college course in biology or biological anthropology. As such, the following information is not meant to be a detailed discussion of genetics, but instead a brief review of some high points and terminology in order to dive into population genetics as quickly as possible. More detail will be given as needed throughout the text. If you find that the following brief review is a bit too brief, I suggest getting more review and/or detail from comprehensive Internet sources such as *Wikipedia*, browsing through some introductory genetics books, and consulting with your professor.

Most discussions of genetics start with mention of deoxyribonucleic acid (DNA), often referred to casually as ''the genetic code.'' Although we are learning more every day about the nature of DNA and how it works, many of the basic principles of population genetics were derived long before much was known about DNA. Indeed, James Watson and Francis Crick discovered the biochemical structure of DNA in 1953, whereas many ideas in population genetics were first developed in the 1930s and 1940s. Although advances in molecular genetics have certainly affected continued development of population genetics in terms of both theory and methods (as will be described later), many of the basic concepts of genetic transmission in populations were developed before we really knew the structure and function of exactly what was being transmitted.

The DNA molecule is made up of two strands that consist of nucleotides, molecules that contain a nitrogen base connected to sugar and phosphate groups. There are four different bases in DNA: adenine (A), thymine (T), cytosine (C), and guanine (G). The sequence of these four different bases make up the genetic ''code,'' and by analogy they can be considered ''letters'' in a four-letter DNA alphabet. A related molecule, ribonucleic acid (RNA), is involved in the transcription of proteins, expression of genes, and other vital biochemical functions. A critical aspect of DNA is that the A and T bases pair up as do the C and G bases. As DNA is double-stranded, this means that an A on one strand is paired with a T on the other strand. Likewise, T is paired with A, C with G, and G with C. This property of DNA allows it to make copies of itself, thus ensuring the transmission of genetic information from cell to cell. The pairing of bases between the two stands is known as a **base pair** (abbreviated bp), and the length of DNA sequences is measured by the number of base pairs.

A. Mendel's Laws

Much (though not all) of our DNA exists on long strands in the nuclei of our cells, called **chromosomes**. Chromosomes come in pairs. Different species have different numbers of chromosomes; humans have 23 pairs, whereas chimpanzees (our closest living relative) have 24 pairs. During the replication of body cells through **mitosis**, a single cell containing 23 pairs of chromosomes will duplicate, giving rise to two identical cells, each with 23 pairs of chromosomes. However,

this is not what happens during reproduction. Instead of passing along 23 pairs of chromosomes to your offspring in a sex cell (sperm in males, egg in females), you pass on *one* of each pair through the process of **meiosis**. The process of chromosome pairs separating through meiosis is also known as **Mendel's law of segregation** (or, sometimes, as Mendel's first law). You contribute 23 chromosomes (but not 23 *pairs*), and your mate contributes 23 chromosomes, resulting in your child having $23 + 23 = 23$ chromosome pairs. Likewise, your genetic inheritance also resulted from this process, as one of each chromosome pair came from your mother and the other one came from your father.

As a bisexual organism (a species that has two distinct sexes, male and female), half of your genetic inheritance comes from your mother and half from your father. The same applies to any biological siblings. Apart from identical twins, why are you not genetically identical to a sibling? If my brother and I both received 50% of our DNA from our mother and 50% from our father, why are we not genetically the same? The answer relates to basic probability; we do not inherit the *same* 50%. For any given chromosome pair, there is a 50 : 50 chance of one being passed on to an offspring, either the maternal chromosome (from your mother) or the paternal chromosome (from your father). For example, imagine that I have passed along my maternal chromosome for the first chromosome pair to a child. The next child may or may not receive the same maternal chromosome; it is a 50 : 50 chance for either the maternal or the paternal chromosome. The same probability applies to each chromosome pair, as they are all *independent* such that whatever chromosome you pass on from the first chromosome pair has no effect on the second pair, the third, and so on.

We can illustrate this principle with a simple analogy using coins. Imagine an organism with only three chromosome pairs, each represented by a penny with two sides—heads and tails. If we flip the first coin, we have a 50 : 50 chance of getting heads (H) or tails (T). We will use this as a model for a chromosome pair consisting of one chromosome labeled H and one labeled T. If you flip heads for the first coin (chromosome pair), what is the probability of flipping heads on the *second* coin? It is still 50 : 50 because the coin flips are independent; the outcome of one coin flip does not influence any other coin flips. In terms of the genetic analogy, this hypothetical organism can produce eight different combinations of coin flips. One of these eight combinations would be getting heads for the first coin, heads for the second coin, and heads for the third coin. Another possibility would be heads for the first coin, heads for the second coin, and tails for the third coin. If we follow this pattern, we wind up with eight different combinations, each equally likely:

1. Heads–heads–heads
2. Heads–heads–tails
3. Heads–tails–heads
4. Heads–tails–tails
5. Tails–heads–heads
6. Tails–heads–tails
7. Tails–tails–heads
8. Tails–tails–tails

Because of chance, this organism could produce eight different combinations of chromosomes. This independent inheritance is known as **Mendel's law of independent assortment** (or Mendel's second law).

In principle, we could simulate the same process for human beings by using 23 different coins, but it take much too long to enumerate all possible combinations of coin flips. Instead, we can figure out the number of possibilities using the simple formula 2^n, where n is the number of coins/chromosome pairs. For humans, $n = 23$ chromosome pairs, giving $2^{23} = 8,333,608$ combinations! Keep in mind that this is for one individual. The same rule applies to the production of sex cells in the individual's mate; they, too, can produce up to 8,388,608 combinations. A child could therefore have any of the first parent's combinations paired with any of the second parent's combinations, giving a total of $8,388,608 \times 8,388,608 = 70,368,744,177,664$ possible genetic combinations in any given child! Given the number of possibilities, it is easy to see why it would be virtually impossible for me to be genetically identical to my nontwin brother for my entire **genome**.

As is typically the case when explaining basic models of reality, I have to point out that all of the above is actually a bit of an oversimplification. The basic process is further complicated by **recombination**, which involves the crossover of sections of DNA of chromosome pairs during meiosis. Start with a pair of chromosomes, with one chromosome from the mother and one from the father. During meiosis, the pair does no segregate exactly, such that pieces of the mother's DNA are exchanged with pieces of the father's DNA. Thus, any sex cell that you pass on to an offspring is unlikely to follow the ideal Mendelian model of being either your mother's chromosome or your father's chromosome, but instead reflects parts of both. The process of recombination provides even more shuffling of genetic combinations with each generation.

Through meiosis with recombination, a new generation can reflect different combinations of what was present in the parental generation. However, in terms of the overall genetic composition of the population (how many different genetic forms exist), this reshuffling does not change anything. An analogy here would be a deck of cards. Each time you shuffle the deck and deal out a five-card poker hand, you are likely to get a different combination, such as a three of clubs, five of spades, six of spades, ten of hearts, and a queen of diamonds. Return these cards to the deck, shuffle, and deal again. You are most likely to have a completely different hand (it is possible to get the same hand, but extremely unlikely, as there are 2,598,960 possible different five-card poker hands using 52 cards and no jokers). Each time you shuffle and deal, you can get a new combination, but the basic composition of the deck has not changed—you still have four suits each with 13 cards ranging from 2 through ace. Nothing new would happen unless there were a *mutation* in the deck, say, resulting from changing a 10 of spades to a brand new type of card, such as an 11 of spades. (Don't try this in a real game!) Population genetics involves understanding how the genetic composition of a population can change through the operation of mutation and other forces of evolution.

B. Alleles, Genotypes, and Phenotypes

What is a **gene**? As with many core ideas and concepts (e.g., life, love, culture, race), the actual definition of gene has changed over time and is often difficult to pin down (Marks and Lyles 1994). The term *gene* was first used in a very general way to refer to a *unit of inheritance*. With the growth of molecular genetics, it has become more common to refer to a gene in a more specific sense, which is a DNA sequence associated with a functional product, such as a protein. This more restricted definition does not include noncoding sections of DNA. Although some population geneticists use the more current restricted definition (e.g., Hamilton 2009), others use the more general definition for convenience (e.g., Hedrick 2005). Here, I will use the more specific restricted definition to comply with your likely background in genetics, and refer to the entire genome as consisting of genes and other DNA sequences. The broader term **genetic marker** is often used to refer to any gene or DNA sequence that has a known location on a specific chromosome.

When we study a genetic marker, we refer to its specific location on a particular chromosome; this location is referred to as a **locus** (plural **loci**). A key concept in population genetics is the **allele**, which refers to alternative forms of a gene or DNA sequence at a given locus. Loci that have two or more alleles that are not rare (typically defined as a frequency greater than 0.01) are called **polymorphisms**, which literally translates as ''many forms.''

As an example of the concept of allele, consider the gene that affects lactase production in humans. As mammals, humans rely on milk during infancy. We produce an enzyme (lactase) in order to break down milk sugar (lactose). A specific gene (LCT) is located on chromosome 2 and regulates the production of lactase. There are several different forms (alleles) of this gene. One allele (*R*) causes enzyme production to decrease during early childhood (an age by which humans have been weaned), and another allele (*P*) allows continued high production of lactase into adulthood, a condition known as *lactase persistence*. There is also a third rare allele, but it will not be discussed in this example (Mielke et al. 2011).

For any trait in your nuclear DNA, including lactase activity, you inherit two copies of the gene or DNA sequence, one from your mother and one from your father, which collectively makes up your **genotype**. In the case of lactase activity, there are two main alleles (*R* and *P*) in the human species, which means that there are three possible genotypes. Some individuals will inherit two copies of the *P* allele and will have the genotype *PP*, while others will inherit two copies of the *R* allele and have the *RR* genotype. Both people with *PP* and *RR* genotypes are **homozygous** for this trait, which means that they have inherited the same allele from both mother and father. There is a third possibility, which is the genotype *PR*, where the person has inherited a *P* allele from one parent and an *R* allele from another parent (it does not matter which parent gave the *P* allele and which gave the *R* allele). When someone inherits a different allele from each parent, that person is **heterozygous** for that trait.

What are the different outcomes for these different genotypes? Each has inherited genetic information regarding the restriction or persistence of lactase production. The physical manifestation of a genotype is known as the **phenotype**. In complex traits, such as height or skin color, the phenotype is a reflection

of the genotypes of the different genes that affect the trait as well as environmental effects, such as nutrition in the case of height, or solar exposure in the case of skin color. In ''simple'' genetic traits, such as lactase activity, the phenotype is determined by the genotype and which, if any, alleles are dominant or recessive. The effect of a **dominant** allele is noticeable even if only one copy is present, whereas a **recessive** allele's effect can be masked by a dominant allele. In the case of lactase activity, the *P* allele (lactase persistence) is dominant and the *R* allele (restriction) is recessive. This means that someone who inherits one *or* two *P* alleles will show the lactase persistence phenotype, and those that inherit two *R* alleles will show lactase restriction. In other words, lactase persistence can result from either the *PP* or *PR* genotypes, and lactase restriction can result only from having the *RR* genotype.

It is important to remember that *dominant* and *recessive* refer to the nature of the alleles and have nothing to do with the actual frequency of an allele; that is, a dominant allele is not necessarily more common than a recessive allele. For example, in humans there is a condition resulting in extra fingers or toes (polydactyly) that is caused by a dominant allele, yet it is very rare in occurrence (Wolf and Myrianthopoulos 1973). Another example in humans is the ABO blood group, where the most common allele in our species is the *O* allele, which is recessive.

For any given locus, the alleles need not be either dominant or recessive. For many loci, the alleles are **codominant**, meaning that the effect of both alleles is expressed in the phenotype. An example in humans is the MN blood group located on chromosome 4, which has two alleles, *M* and *N*, which produce different molecules on the surface of our red blood cells; the *M* allele produces type M molecules and the *N* allele produces type N molecules. Given these two alleles, we have three possible genotypes: *MM, MN*, and *NN*. What about the phenotypes? Logically, we can see that the homozygous genotype *MM* will result in type M molecules because both alleles contain the same message—type M blood. It is also clear that the genotype *NN* will produce type N molecules. What of the genotype *MN*? The phenotype associated with a heterozygous genotype depends on whether one allele was dominant. In this case, the *M* and *N* alleles are codominant, which means that *both* the *M* allele and the *N* allele will manifest, resulting in the production of *both* type M and type N molecules. In the case of a codominant locus, each genotype has a distinct phenotype. As we will see in Chapter 2, this makes it much easier to count alleles and determine their frequency (a vital part of population genetics).

Before moving on, I want to point out some other complications. Although most examples in this book use a simple model of a single locus with two alleles, in reality there are actually many loci with more than two alleles, and some loci where there are dozens of alleles. Basic concepts will be introduced using the simple two-allele model where possible and bringing in this additional complication where appropriate.

Another complication is the fact that some loci have dominant, recessive, *and* codominant alleles. A good example for humans is the ABO blood group, located on chromosome 9. There are three main alleles, *A, B*, and *O*, where the *A* allele codes for type A molecules, the *B* allele codes for type B molecules, and the *O* allele

codes for neither of these. In the ABO system, the *O* allele is recessive and the *A* and *B* alleles are codominant. Given three possible alleles, there are six possible genotypes: *AA*, *BB*, *OO*, *AO*, *BO*, and *AB*. What are the possible phenotypes?

The phenotypes of the three homozygous genotypes (*AA, BB, and OO*) are easy to determine. Genotype *AA* produces type A blood, genotype *BB* produces type B blood, and genotype *OO* produces type O blood. The phenotypes of the three remaining genotypes can be determined by knowing which alleles are dominant, recessive, or codominant. Because the *O* allele is recessive, those with genotype *AO* will show only the effect of the dominant *A* allele, and hence will have type A blood. Likewise, those with genotype *BO* will show only the effect of the dominant *B* allele, and will have type B blood. The remaining genotype, *AB*, has two codominant alleles, which means that both A and B molecules will be produced, and people with this genotype therefore have what we call type AB blood. For the ABO blood group, there are three alleles that can form six different genotypes that correspond to four different phenotypes (ignoring for the moment additional complications, such as the fact that there are actually two subtypes of the *A* allele).

C. How Do We Assess Human Genetic Diversity?

As will be clear in later chapters, much of the core of population genetics theory is abstract, dealing with hypothetical alleles at hypothetical loci in hypothetical populations. Although hypothetical rumination is interesting in and of itself, the ultimate test of a mathematical model of reality is to see how well it represents reality, which means that at some point we need information about real alleles at real loci in real populations! Although a variety of loci and traits will be provided in case studies throughout this book, it is useful to look briefly at some of the different ways anthropologists and geneticists use to assess genetic diversity.

Red Blood Cell Markers

For the first half of the twentieth century, most information on genetic diversity in human populations came from the study of blood types based on red blood cell groups, where phenotypes were based on the reaction of antigens present in the blood with corresponding antibodies (Boyd 1950). In the ABO blood group system, for example, this is based on reactions of A and B antigens with their respective antibodies, anti-A and anti-B. Suppose that someone's blood shows a reaction with the anti-A antibody but not the anti-B antibody. This means that they have the A antigen but not the B antigen, and therefore have blood type A, and therefore either the *AA* or the *AO* genotype. There are many different red blood cell systems, including ABO, Rhesus, MN, Kell, Diego, Duffy, and P (Crawford 1973).

By the 1960s and 1970s, technological advances such as electrophoresis had led to a proliferation of other genetic markers of the blood. **Electrophoresis** involves passing an electric current through a gel. Blood samples are placed at the negative pole of the gel and, as current flows from negative to positive, molecules move through the gel. Because different molecular structures move at different rates, the process allows identification of different molecular structures associated with different genotypes. Applied to blood samples from anthropological surveys, a vast

amount of data were collected on numerous red blood cell protein and enzyme loci (Crawford 1973; Roychoudhury and Nei 1988; Cavalli-Sforza et al. 1994).

DNA Markers

Genetic markers of the blood, including markers based on white blood cells, are now labeled as *classical genetic markers*, contrasted with the newer DNA markers. Although classical markers provide information on genetic variation, DNA markers provide a closer window on genetic variation, moving beyond the level of molecular variability to the underlying level of DNA variation.

One method of DNA analysis involves the identification of **restriction fragment length polymorphisms** (RFLPs). Restriction enzymes that are produced by different types of bacteria can bind to sections of a DNA sequence and cut that sequence at a particular point. For example, the *EcoRI* bacterial enzyme will bind to the 6-bp sequence GAATCC (which, by definition, corresponds to the sequence CTTAGG on the other DNA strand). If this sequence is present in a DNA sample, *EcoRI* will cut the sequence between the G and the first A, producing two fragments, one with the base G and the other with the sequence AATCC. If the DNA sample did not contain the sequence GAATCC, but instead had a mutation resulting in GATTCC (where the second A mutated into T), then the target sequence would not be recognized and the DNA sample would not be cut. Depending on the presence or absence of certain DNA sequences, a DNA sample might be cut into fragments of different lengths.

Another type of DNA variation widely studied in human populations consists of repeated DNA sequences, such as CACACACACACACA, where the 2-bp sequence is repeated 7 times. Because of mutation, the number of repeats can go up or down, resulting in variation. **Short tandem repeats** (STRs), also known as *microsatellite DNA*, are widely used in studies of human populations. STRs consist of short repeated sequences consisting of 2–5 bp. Longer repeated sequences, known as *minisatellites*, are also used.

Another form of DNA analysis looks for **single-nucleotide polymorphisms** (SNPs), where DNA sequences differ by one base, such as having the base C on one sequence versus the base T on another sequence:

Sequence 1: TATTCCGGA
Sequence 2: TACTCCGGA

In this case, the two sequences differ at the third position, and there are two alleles: the first has the base T and the second has the base C. SNP variation is being increasingly studied in human populations; as of 2007, over 3.1 million SNPs had been identified (International HapMap Consortium 2007).

Haplotypes

Loci that are close together on the same chromosome tend to be inherited together (**linkage**). A **haplotype** is a combination of alleles that are inherited as a single unit. Haplotypes are sometimes defined as a set of linked loci, and can be based on RFLPs, STRs, SNPs, or combinations of these. For example, Foster et al. (1998)

conducted a genetic analysis on descendants of male relatives of the third US president, Thomas Jefferson, to see if there was a genetic connection between this family and the descendants of Eston Hemings, son of Sally Hemings, who was an enslaved African-American woman. This study involved using a haplotype of the Y chromosome that was unique to the Jefferson family. This haplotype consisted of seven SNPs, 11 STRs, and 1 minisatellite.

Mitochondrial DNA and Y-Chromosome DNA

The discussion so far has dealt with nuclear DNA and inheritance from both parents, the traditional way to present Mendelian genetics. These examples refer to **diploid** inheritance (two copies). Although the majority of examples presented in this text refer to diploid inheritance, some traits in humans are **haploid**, and come from only one parent. One type of haploid inheritance is **mitochondrial DNA** (abbreviated mtDNA). Although most of our DNA is contained in the chromosomes, there is a small amount in the mitochondria, which are the cell structures responsible for energy production. In humans, mitochondrial DNA, the circular DNA molecule is typically 16,569 base pairs in length, which is a very small fraction of the more than 3 billion base pairs of nuclear DNA.

What makes mitochondrial DNA so fascinating and useful is the way it is inherited. Unlike nuclear DNA, which is inherited from both parents, you inherit mitochondrial DNA only from your mother. This pattern of inheritance results from the way in which sperm and egg combine to form a zygote (fertilized egg); the mitochondria in the zygote comes from the egg, and thus contains only the mother's genetic contribution. Transmission from one generation to the next is through the female line. Although males inherit mitochondrial DNA from their mothers, they cannot pass it on; transmission occurs only in the female line.

This exclusive maternal inheritance simplifies genealogical analysis, figuring out where certain alleles came from. With nuclear DNA, it is difficult to do this because the number of potential ancestors doubles with each generation in the past—you have two parents, four grandparents, eight great grandparents, and so on. The number of potential ancestors and recombination every generation make it difficult to tell if a particular allele came from any given ancestor. With mitochondrial DNA, you have only one ancestor in any generation in the past. One generation back, that ancestor would be your mother; two generations back would be your mother's mother, and so on into the past. Another advantage of mitochondrial DNA is its use in ancient DNA analysis because the high number of copies per cell means it that will be more likely to survive degradation compared with nuclear DNA (O'Rourke 2007). Mitochondrial DNA is also used in ancestry testing, a service available from a number of vendors. The problem here is that such tests can tell you about ancestry in only one line. For example, you may have a maximum of 16 great-great grandparents (see Chapter 3 for why this is a maximum and how you can actually have fewer great-great grandparents). Mitochondrial DNA analysis will tell you about only one of these 16 ancestors—your mother's mother's mother's mother—and not your mother's mother's mother's father, your mother's mother's father's mother, or any of your other ancestors.

Just as ancestry can be traced in the maternal line using mitochondrial DNA, a similar pattern of ancestry is found in **Y-chromosome DNA**. Of our 23 pairs of chromosomes, one pair is the sex chromosomes and the other 22 pairs are referred to as **autosomes**. There are two types of sex chromosomes, X and Y, which determine biological sex; females have two X chromosomes, whereas males have one X chromosome and one Y chromosome. Males receive their X chromosome from their mother and their Y chromosome from their father. When sperm are produced, the sex chromosomes segregate and males pass on either their X or their Y. Because there is very little recombination of the Y chromosome with the X chromosome, this means that males pass on their Y chromosome almost intact to their sons. Analysis of the nonrecombining part of the Y chromosome provides the same insights as mitochondrial DNA, although in the father's line.

Quantitative Traits

Although the focus of this book is primarily on the application of population genetics theory to a simple single-locus model, it is useful to point out that there are also extensions to more complex traits, particularly those that involved the effects of both genetics and the environment. Before the dawn of the twentieth century, the only way to approximate genetic variation was through physical traits, such as cranial and facial measurements. The study of **quantitative genetics** deals with such traits. Many physical traits, such as height, head length, and skin color, are examples of **quantitative traits**, whose phenotype varies continuously. For example, if we consider height, and see one person with a height of 5 ft 6 in. and another person with a height of 5 ft 7 in., we know that it is possible to find someone with any intermediate value, such as 5 ft 6.5 in., 5 ft, 6.6 in., and so on.

Quantitative traits are due to the joint influence of one or more loci and environmental influences, where the latter can include a variety of influences ranging from prenatal environment to climate to diet, among many others. Often, quantitative traits are **polygenic**, meaning that two or more loci interact to produce a genotype. In some cases, an **equal and additive effects model** can be used to describe polygenic inheritance, where all of several loci contribute equally to the same phenotype. In other cases, one locus may have a more substantial effect, such as in a **major gene model**. In all cases, the phenotype (e.g., how tall you are, the color of your skin) reflects the joint effect of these loci and environmental influences. Thus, two people could have the same genotypic inheritance but grow to different heights because of exposure to different environmental conditions. Likewise, two people may have different genotypes but wind up with the same phenotype because of environmental influences.

With the advent of biochemical and DNA markers, attention has moved away from quantitative traits. This does not mean, however, that they are without value. Although newer analytic methods have now allowed such traits to provide useful measures of human population relationships and history (Relethford 2007; von Cramon-Taubadel and Weaver 2009), they will only be referenced briefly in this text, with the bulk of attention given to ''simpler'' genetic traits.

III. PRINCIPLES OF PROBABILITY

Population genetics is mathematical in nature. The use of mathematics can be scary to many students. Indeed, I find some of the more complex methods in the field scary myself! However, the basic concepts of population genetics can be learned with a minimum of mathematics. The level of math used in this text assumes some previous background in the basic algebra learned in high school. I also use many graphs to give a visual feel for the mathematical relationships that may not be directly apparent from the raw equations. Some algebra will be used for proofs, many of which are explained at the end of each chapter to avoid interference with the flow of text. There will be one placed where additional explanation will be appended for those who know a bit of calculus, but that is not necessary for understanding the basic concepts.

Most of what we look at in population genetics to model the evolutionary process consists in applying some basic concepts of probability. If you think about it, a number of questions about reproduction and inheritance boil down to questions about probability. For example, what is the probability that a parent will pass on a given allele to a child? At the population level, similar questions regarding the probability of genetic transmission apply. What is the probability that a given locus will mutate within a generation? What is the probability that a given allele will increase over time, decrease over time, or stay the same? What is the probability that an allele will move from one population to another through migration in a given generation? These and other questions boil down to questions of probability.

A. Some Simple Rules of Probability

Many basic concepts of probability can be demonstrated using coins, dice, and/or a deck of cards. Such everyday objects are useful in learning more abstract concepts because the answers are often more intuitive. For example, consider the very simple questions that can be generated using a single coin; for instance, what is the probability of flipping the coin and getting heads? We all can answer this question immediately—the probability is $\frac{1}{2}$. The same is true for the probability of flipping the coin and getting tails ($= \frac{1}{2}$). This is probably obvious, but where exactly did we get this number? We simply divided the number of times that a specific event (getting a head) could occur by the total number of events that could occur (getting a head or a tail). Consider a different example. Roll a single die, a cube with six numbered sides. What is the probability of rolling the die and getting the number 3? There is only one way to get the number 3, and there are six possible outcomes, so the probability is $\frac{1}{6}$. The same is true of all the other numbers (1, 2, 4, 5, 6); each has a $\frac{1}{6}$ probability. The sum of the probabilities of all possible outcomes is equal to 1. In the case of flipping a coin, there are only two possible outcomes (heads and tails), each with a $\frac{1}{2}$ probability, and the sum of all outcomes is $\frac{1}{2} + \frac{1}{2} = 1$. For rolling a single die, there are six possible outcomes, each with $\frac{1}{6}$ probability, and they add up to 1.

Probabilities are expressed as proportions that can range from 0 to 1. For example, $\frac{1}{6} = 0.167$. Sometimes we here probabilities also expressed as

percentages, and all you have to remember is that a percentage is a proportion multiplied by 100. For example, we can say that the probability of rolling a single die and getting the number 3 is $\frac{1}{6} = 0.167$ (a proportion), and we could express this as a percentage (16.7% of the time, we will get the number 3).

In population genetics, we often look at events that are independent of one another, such as Mendel's law of independent assortment. For example, what is the probability of rolling a pair of dice and getting the number 3 on both dice? Again, most of us will tend to solve such problems intuitively. We know the probability of rolling a die and getting the number 3 is $\frac{1}{6}$, and we also know that what comes up on one die in no way influences what comes up on the second die (they are independent). Thus, the probability of getting the number 3 on both dice is $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0.028$. What we have done here is use what is known as the AND rule in probability. In general, we use the symbols $P(A)$ to refer to the probability that outcome A occurs and the symbol $P(B)$ to refer to the probability that outcome B occurs, such that the probability of both A AND B occurring is their product:

$$P(A \text{ AND } B) = P(A)P(B) \tag{1.1}$$

In formal terms, the question above of rolling a pair of dice and getting two 3s is solved using the AND rule as $P = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0.028$. As an aside, you might notice that this equation is numbered. All main equations in this text are numbered, consisting of a chapter number (the first number), followed by a period, and followed by a sequence number. Thus, equation (1.1) refers to the first numbered equation in Chapter 1. Main equations are those that are referenced elsewhere in the text.

We are also interested in situations where two outcomes are mutually exclusive, and therefore both cannot be true at the same time. For example, what is the probability of rolling a single die and getting a 3 OR a 4? The probability of rolling a 3 is $\frac{1}{6}$, as is the probability of rolling a 4. These outcomes are mutually exclusive, as you cannot roll a 3 and a 4 at the same time! What then is the probability of A *or* B? The answer uses what is known as the OR rule, where the individual probabilities are added:

$$P(A \textit{ or } B) = P(A) + P(B) \tag{1.2}$$

Thus, the probability of rolling a die and getting a 3 or a 4 is $\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0.333$.

Here are a few simple examples for reviewing the basic rules of probability:

1. ***Question***: You roll a pair of dice. What is the probability of getting an even number?
   Answer: There are three possible even numbers (2, 4, and 6) out of six possible outcomes. The probability is $\frac{3}{6} = \frac{1}{2} = 0.5$.
2. ***Question***: You flip two coins at the same time. What is the probability of having both of the coins come up heads?
   Answer: This is a case that calls for the AND rule, as the answer is the product of each coin coming up heads. The probability is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = 0.25$.

3. ***Question***: This is the same problem as 2, but this time, flip *three* coins. What is the probability of all three coming up heads?
   Answer: You still use the AND rule but extend it to three outcomes by multiplying the three probabilities as $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8} = 0.125$.
4. ***Question***: You have thoroughly shuffled a standard 52-card deck of cards (no jokers!). What is the probability of randomly selecting a card and getting an ace or a face card?
   Answer: You need to use the OR rule to compute the probability. Start with the probability of getting an ace. There are four aces in the deck, so the probability of drawing an ace is $\frac{4}{52}$. There are nine face cards in the deck (three jacks, three queens, and three kings). The probability of getting a face card is therefore $\frac{9}{52}$. Put these together using the OR rule, and the probability of getting an ace or a face card is $\frac{4}{52} + \frac{9}{52} = \frac{13}{52} = \frac{1}{4} = 0.25$.

B. Genetics and Probability

Many problems in Mendelian genetics concern probability. A typical question asks the student to describe the distribution of possible genotypes and phenotypes in the offspring of a given mating. To illustrate this, consider a hypothetical locus that has two alleles, *A* and *a*. With two alleles, there are three possible genotypes: *AA*, *Aa*, and *aa*. What is the probability of two *Aa* parents having an offspring with the genotype *AA*? To answer this, we have to connect the basic idea of inheritance with probability. A child who has the *AA* genotype will have inherited the *A* allele from both parents. The question can now be expressed more specifically: What is the probability that both parents pass on an *A* allele? Both parents have the genotype *Aa*, which means that each parent has a $\frac{1}{2}$ chance of passing on the *A* allele. Given this, we use the 'and' rule and give the answer as $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = 0.25$.

What about when the parents have different genotypes? The same principles apply. For example, what is the probability of a man with genotype *AA* and a woman with genotype *Aa* having a child with genotype *Aa*? The child must inherit an *A* allele from one parent and an *a* allele from the other parent. In this case, the father can pass on only the *A* allele because he has the *AA* genotype (and obviously cannot pass on a different allele). Therefore, the mother must pass on the *a* allele. What is the probability of the man passing on an *A* allele *and* the woman passing on the *a* allele? Because the man only has *A* alleles, the probability that he will pass on an *A* allele is $\frac{2}{2} = 1$. The woman has the *Aa* genotype, and therefore the probability of her passing on the *a* allele is $\frac{1}{2}$. Using the AND rule, we get the answer to the original question as $1 \times \frac{1}{2} = \frac{1}{2} = 0.5$. In other words, we expect that half the time this couple will have a child with the *Aa* genotype.

It is often more useful to consider the distribution of all possible outcomes of a given mating. A useful tool is the **Punnett square**, a simple method invented by geneticist Reginald Punnett (1875–1967) and one that you may recall from high school and/or college biology. In a Punnett square, we simply construct a 2×2 table that lists the possible contributions of each parent. For the above example, where the father has the *AA* genotype and the mother has the *Aa* genotype, the Punnett square is as follows: