POPULATION GENETICS AND MICROEVOLUTIONARY THEORY

Alan R. Templeton

Department of Biology
Washington University
St. Louis, Missouri

WILEY-LISS
A John Wiley & Sons., Inc., Publication

POPULATION GENETICS AND MICROEVOLUTIONARY THEORY

POPULATION GENETICS AND MICROEVOLUTIONARY THEORY

Alan R. Templeton

Department of Biology
Washington University
St. Louis, Missouri

WILEY-LISS
A John Wiley & Sons., Inc., Publication

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

Library of Congress Cataloging-in-Publication Data:

Templeton, Alan Robert.
Population genetics and microevolutionary theory/ Alan R. Templeton
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-471-40951-9 (cloth)
ISBN-10: 0-471-40951-0 (cloth)
1. Population genetics. 2. Evolution (Biology) I. Title.

QH455.T46 2006
576.5′8–dc22 2006042030

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Bonnie
and to the Memory of Hampton Carson

CONTENTS

PREFACE

I have been teaching population genetics for 30 years, and during that time the importance and centrality of this field to modern biology have increased dramatically. Population genetics has always played a central role in evolutionary biology as it deals with the mechanisms by which evolution occurs within populations and species, the ultimate basis of all evolutionary change. However, as molecular genetics matured into genomics, population genetics was transformed from a discipline receiving new techniques from molecular genetics into a discipline providing the basic analytical methods for many aspects of genomics. Moreover, an increasing number of students are interested in the problems of species extinction and of environmental degradation and change. Population genetics offers many basic tools for conservation biology as well. As a result, the audience for population genetics has increased substantially, and I have witnessed a sixfold increase in the enrollment in my population genetics course over the past several years. This book is written with this expanded audience in mind. Many examples are given from conservation biology, human genetics, and genetic epidemiology, yet the focus of this book remains on the basic microevolutionary mechanisms and how they interact to create evolutionary change. This book is intended to provide a solid basis in population genetics both for those students primarily interested in evolutionary biology and genetics as well as for those students primarily interested in applying the tools of population genetics, particularly in the areas of conservation biology, human genetics, and genomics. Without a solid foundation in population genetics, the analytical tools emerging from population genetics will frequently be misapplied and incorrect interpretations can be made. This book is designed to provide that foundation both for future population and evolutionary geneticists and for those who will be applying population genetic concepts and techniques to other areas.

One theme throughout this book is that many important biological phenomena emerge from the interactions of two or more factors. As a consequence, evolution must be viewed with a multidimensional perspective, and it is insufficient to examine each evolutionary force one by one. Two highly influential mentors strengthened this theme in my work: Charles Sing and Hampton Carson. Charlie was my Ph.D. advisor and continues to be a mentor, collaborator, and friend. Charlie always stressed the importance of interactions in biology and genetics, and he was and is concerned with the "big picture" questions. I cannot thank Charlie enough for his continuing intellectual challenges and for his friendship.

Hamp Carson also stressed the importance of interacting forces in evolution and genetics. Hamp was both my undergraduate research mentor and my postdoctoral advisor, as well as a long-time collaborator and friend. Hamp died at the age of 91 as this book was nearing completion. He lived a full and highly productive life, and I dedicate this book in his memory to honor his life and accomplishments.

Many of my graduate students, both current and former, contributed significantly to this book. Indeed, the impetus for writing this book came largely from two former graduate students, Delbert Hutchison and Keri Shingleton. When Delbert and Keri were at Washington University as graduate students, they also served as teaching assistants in my population genetics course. My lectures did not follow any of the existing textbooks, so first Delbert, and then Keri, wrote out detailed lecture notes to help the students. These notes also formed the backbone of this book, and both Delbert and Keri strongly urged me to take their notes and transform them into a book. This is the book that resulted from that transformation.

Many of my graduate students read drafts of the chapters and offered many suggestions that were incorporated into the book. I thank the following graduate students for their valuable input: Corey Anderson, Jennifer Brisson, Nicholas Griffin, Jon Hess, Keoni Kauwe, Rosemarie Koch, Melissa Kramer, Taylor Maxwell, Jennifer Neuwald, James Robertson, and Jared Strasburg. In addition, many of my former graduate students and colleagues read drafts of this book and often used these drafts in teaching their own courses in population genetics. They also provided me with excellent feedback, both from themselves and from their students, so I wish to thank Reinaldo Alves de Brito, Keith Crandall, Delbert Hutchison, J. Spencer Johnston, and Eric Routman. I also want to thank three anonymous reviewers for their comments and suggestions on the first six chapters of this book. Finally, I used drafts of this book as my text in my population genetics class at Washington University. Many of the students in this class, both graduate and undergraduate, provided me with valuable feedback, and I thank them for their help.

1

SCOPE AND BASIC PREMISES OF POPULATION GENETICS

Population genetics is concerned with the origin, amount, and distribution of genetic variation present in populations of organisms and the fate of this variation through space and time. The kinds of populations that will be the primary focus of this book are populations of sexually reproducing diploid organisms, and the fate of genetic variation in such populations will be examined at or below the species level. Variation in genes through space and time constitute the fundamental basis of evolutionary change; indeed, in its most basic sense, **evolution** is the genetic transformation of reproducing populations over space and time. Population genetics is therefore at the very heart of evolutionary biology and can be thought of as the science of the mechanisms responsible for **microevolution,** evolution within species. Many of these mechanisms have a great impact on the origin of new species and on evolution above the species level (macroevolution), but these topics will not be dealt with in this book.

BASIC PREMISES OF POPULATION GENETICS

Microevolutionary mechanisms work upon genetic variability, so it is not surprising that the fundamental premises that underlie population genetic theory and practice all deal with various properties of deoxyribonucleic acid (DNA), the molecule that encodes genetic information in most organisms. [A few organisms use ribonucleic acid (RNA) as their genetic material, and the same properties apply to RNA in those cases.] Indeed, the theory of microevolutionary change stems from just three premises:

1. DNA can replicate.
2. DNA can mutate and recombine.
3. Phenotypes emerge from the interaction of DNA and environment.

The implications of each of these premises will now be examined.

Population Genetics and Microevolutionary Theory, By Alan R. Templeton

DNA Can Replicate

Because DNA can replicate, a particular kind of gene (specific set of nucleotides) can be passed on from one generation to the next and can also come to exist as multiple copies in different individuals. Genes therefore have an existence in time and space that transcends the individuals that temporarily bear them. The biological existence of genes over space and time is the *physical basis of evolution*.

The physical manifestation of a gene's continuity over time and through space is a reproducing population of individuals. Individuals have no continuity over space or time; individuals are unique events that live and then die and cannot evolve. But the genes that an individual bears are potentially immortal through DNA replication. For this potential to be realized, the individuals must reproduce. Therefore, to observe evolution it is essential to study a population of reproducing individuals. A reproducing population does have continuity over time as one generation of individuals is replaced by the next. A reproducing population generally consists of many individuals, and these individuals collectively have a distribution over space. Hence, a reproducing population has continuity over time and space and constitutes the physical reality of a gene's continuity over time and space. Evolution is therefore possible only at the level of a reproducing population and not at the level of the individuals contained within the population.

The focus of population genetics must be upon reproducing populations to study microevolution. However, the exact meaning of what is meant by a population is not fixed but rather can vary depending upon the questions being addressed. The population could be a local breeding group of individuals found in close geographic proximity or it could be a collection of local breeding groups distributed over a landscape such that most individuals only have contact with other members of their local group but that on occasion there is some reproductive interchange among local groups. Alternatively, a population could be a group of individuals continuously distributed over a broad geographical area such that individuals at the extremes of the range are unlikely to ever come into contact, or any other grouping of individuals up to and including the entire species. Within this hierarchy of populations found within species, much of population genetics focuses upon the **local population**, or **deme**, a collection of interbreeding individuals of the same species that live in sufficient proximity that they share a system of mating. Systems of mating will be discussed in more detail in subsequent chapters, but for now the **system of mating** refers to the rules by which individuals pair for sexual reproduction. The individuals within a deme share a common system of mating. Because a deme is a breeding population, individuals are continually turning over as births and deaths occur, but the local population is a dynamic entity that can persist through time far longer than the individuals that temporarily comprise it. The local population therefore has the attributes that allow the physical manifestation of the genetic continuity over space and time that follows from the premise that DNA can replicate.

Because our primary interest is on genetic continuity, we will make a useful abstraction from the deme. Associated with every local population of individuals is a corresponding local population of genes called the **gene pool**, the set of genes collectively shared by the individuals of the deme. An alternative and often more useful way of defining the gene pool is that the gene pool is the population of potential gametes produced by all the individuals of the deme. Gametes are the bridges between the generations, so defining a gene pool as a population of potential gametes emphasizes the genetic continuity over time that provides the physical basis for evolution. For empirical studies, the first definition is primarily used; for theory, the second definition is preferred.

The gene pool associated with a local population is described by measuring the numbers and frequencies of the various types of genes or gene combinations in the pool. At this lowest meaningful biological level of a deme, **evolution** is defined as a change through time of the frequencies of various types of genes or gene combinations in the gene pool. This definition is not intended to be an all-encompassing definition of evolution. Rather, it is a narrow and focused definition of evolution that is useful in much of population genetics precisely because of its narrowness. This will therefore be our primary definition of evolution in this book. Since only a local population at the minimum can have a gene pool, only populations can evolve under this definition of evolution, not individuals. Therefore, evolution is an emergent property of reproducing populations of individuals that is not manifested in the individuals themselves. However, there can be higher order assemblages of local populations that can evolve. In many cases, we will consider collections of several local populations that are interconnected by dispersal and reproduction, up to and including the entire species. However, an entire species in some cases could be just a single deme or it could be a collection of many demes with limited reproductive interchange. A species is therefore not a convenient unit of study in population genetics because species status itself does not define the reproductive status that is so critical in population genetic theory. We will always need to specify the type and level of reproducing population that is relevant for the questions being addressed.

DNA Can Mutate and Recombine

Evolution requires change, and change can only occur when alternatives exist. If DNA replication were always 100% accurate, there could be no evolution. A necessary prerequisite for evolution is genetic diversity. The ultimate source of this genetic diversity is mutation. There are many forms of mutation, such as single-nucleotide substitutions, insertions, deletions, transpositions, duplications, and so on. For now, our only concern is that these mutational processes create diversity in the population of genes present in a gene pool. Because of mutation, alternative copies of the same homologous region of DNA in a gene pool will show different states.

Mutation occurs at the molecular level. Although many environmental agents can influence the rate and type of mutation, one of the central tenets of Darwinian evolution is that mutations are random with respect to the needs of the organism in coping with its environment. There have been many experiments addressing this tenet, but one of the more elegant and convincing is replica plating, first used by Joshua and Esther Lederberg (1952) (Figure 1.1). Replica plating and other experiments provide empirical proof that mutation, occurring on DNA at the molecular level, is not being directed to produce a particular phenotypic consequence at the level of an individual interacting with its environment. Therefore, we will regard mutations as being random with respect to the organism's needs in coping with its environment.

Mutation creates allelic diversity. **Alleles** are simply alternative forms of a gene. In some cases genetic surveys focus on a region of DNA that may not be a gene in a classical sense; it may be a DNA region much larger or smaller than a gene or a noncoding region. We will use the term **haplotype** to refer to an alternative form (specific nucleotide sequence) among the homologous copies of a defined DNA region, whether a gene or not. The allelic or haplotypic diversity created by mutation can be greatly amplified by the genetic mechanisms of recombination and diploidy. In much of genetics, recombination refers to meiotic crossing

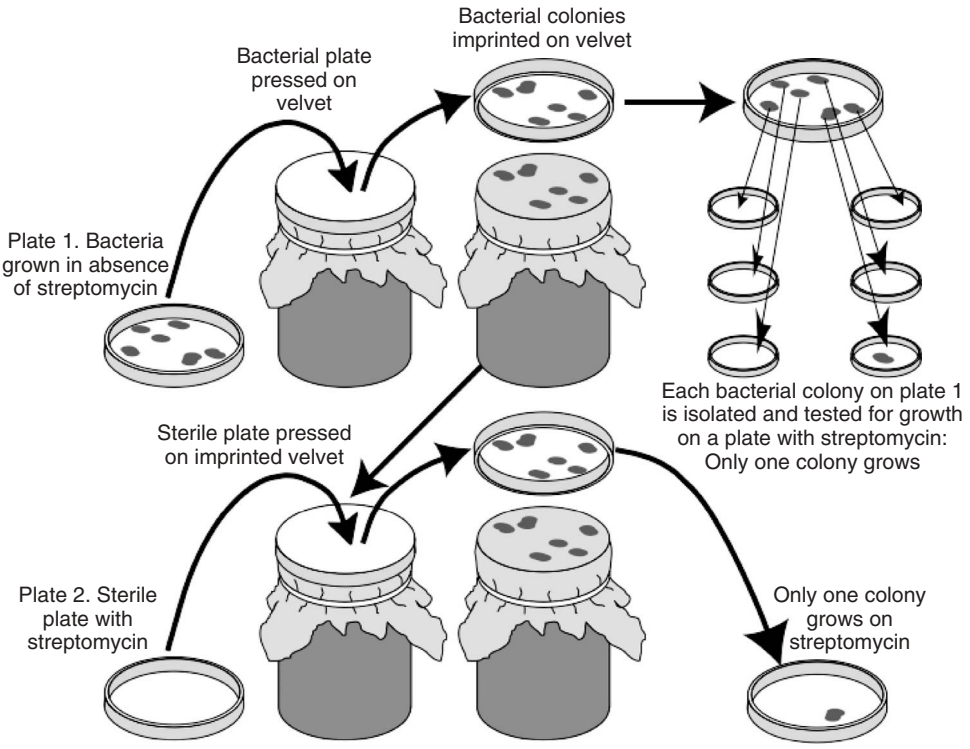

Figure 1.1. Replica plating. A suspension of bacterial cells is spread upon a Petri dish (plate 1) such that each individual bacterium should be well separated from all others. Each bacterium then grows into a colony of genetically identical individuals. Next, a circular block covered with velvet is pressed onto the surface of plate 1. Some bacteria from each colony stick to the velvet, so a duplicate of the original plate is made when the velvet is pressed onto the surface of a second Petri dish (plate 2), called the replica plate. The medium on the replica plate contains streptomycin, an antibiotic that kills most bacteria from the original strain. In the example illustrated, only one bacterial colony on the replica plate can grow on streptomycin, and its position on plate 2 identifies it as the descendant of a particular colony on plate 1. Each bacterial colony on plate 1 is then tested for growth on a plate with the antibiotic streptomycin. If mutations were random and streptomycin simply selected preexisting mutations rather than inducing them, then the colonies on plate 1 that occupied the positions associated with resistant colonies on plate 2 should also show resistance, even though these colonies had not yet been exposed to streptomycin. As shown, this was indeed the case.

over, but we use the term *recombination* in a broader sense as any genetic mechanism that can create new combinations of alleles or haplotypes. This definition of recombination encompasses the meiotic events of both independent assortment and crossing over and also includes gene conversion and any nonmeiotic events that create new gene combinations that can be passed on through a gamete to the next generation. Sexual reproduction and diploidy can also be thought of as mechanisms that create new combinations of genes.

As an illustration of the genetic diversity that can be generated by the joint effects of mutation and recombination, consider the MHC complex (major histocompatibility complex, also known in humans as HLA, human leukocyte antigen) of about 100 genes on the same chromosome. Table 1.1 shows the number of alleles found at 20 of these loci as of 1997 in human populations (Bodmer and Bodmer 1999). As can be seen, mutational changes at these

Table 1.1. Numbers of Alleles Known in 1997 at 20 Loci within Human *MHC* (*HLA*) Region

Locus	Number of Alleles
MHC-1	83
MHC-B	186
MHC-C	42
MHC-E	5
MHC-G	7
MHC-DRA	2
MHC-DRB1	184
MHC-DRB3	11
MHC-DRB4	9
MHC-DRB5	12
MHC-DQA1	18
MHC-DQB1	31
MHC-DOB	1
MHC-DMA	4
MHC-DMB	5
MHC-DNA	1
MHC-DPA1	10
MHC-DPB1	77
TAP1	5
TAP2	4
Total	698

loci have generated from 1 to 186 alleles per locus with a total of 698 alleles over all 20 loci. However, these loci can and do recombine. Hence, recombination has the potential of combining these 698 alleles into 1.71×10^{21} distinct gamete types (obtained by multiplying the allele numbers at each locus). Sexual reproduction has the potential of bringing together all pairs of these gamete types in a diploid individual, resulting in over 10^{42} genotypes and over 10^{33} distinct possible antigenic phenotypes (Bodmer and Bodmer 1999). And this is only from 20 loci in one small region of one chromosome of the human genome! Given that there are only about 6×10^9 humans in the world, everyone on the world (with the exception of identical twins) will have a unique *MHC* genotype when these 20 loci are considered simultaneously. But of course, humans differ at many more loci than just these 20. As of 2004, about 6 million polymorphic nucleotides were known in the human genome. Assuming that most of these are biallelic, each polymorphic nucleotide defines three genotypes, so collectively the number of possible genotypes defined by these known polymorphic sites is $3^{6,000,000} = 10^{2,862,728}$ genotypes. To put this number into perspective, the mass of our entire galaxy in grams is 1.9×10^{44} (Weinberg 1977), a number far smaller than the number of potential genotypes that are possible in humanity just with the known genetic variation. Hence mutation and recombination can generate truly astronomical levels of genetic variation.

The distinction between mutation and recombination is often blurred because recombination can occur within a gene and thereby create new alleles or haplotypes. For example, 71 individuals from three human populations were sequenced for a 9.7-kb region within the *lipoprotein lipase* locus (*LPL*) (Nickerson et al. 1998). This represents just about a third of this one locus. Eighty-eight variable sites were discovered, and 69 of these sites were

used to define 88 distinct haplotypes or alleles. These 88 haplotypes arose from at least 69 mutational events (a minimum of one mutation for each of the 69 variable nucleotide sites) coupled with about 30 recombination per gene conversion events (Templeton et al. 2000a). Thus, intragenic recombination and mutation have together generated 88 haplotypes as inferred using only a subset of the known variable sites in just a third of a single gene in a sample of 142 chromosomes. These 88 haplotypes in turn define 3916 possible genotypes—a number considerably larger than the sample size of 71 people!

Studies such as those mentioned above make it clear that mutation and recombination can generate large amounts of genetic diversity at particular loci or chromosomal regions, but they do not address the question of how much genetic variation is present within species in general. How much genetic variation is present in natural populations was one of the defining questions of population genetics up until the mid-1960s. Before then, most of the techniques used to define genes required genetic variation to exist. For example, many of the early important discoveries in Mendelian genetics were made in the laboratory of Thomas Hunt Morgan during the first few decades of the twentieth century. This laboratory used morphological variation in the fruitfly *Drosophila melanogaster* as its source of material to study. Among the genes identified in this laboratory was the locus that codes for an enzyme in eye pigment biosynthesis known as *vermillion* in *Drosophila*. Morgan and his students could only identify *vermillion* as a genetic locus because they found a mutant that coded for a defective enzyme, thereby producing a fly with bright red eyes. If a gene existed with no allelic diversity at all, it could not even be identified as a locus with the techniques used in Morgan's laboratory. Hence, *all* observable loci had at least two alleles in these studies (the "wildtype" and "mutant" alleles in Morgan's terminology). As a result, even the simple question of how many loci have more than one allele could not be answered directly. This situation changed dramatically in the mid-1960s with the first applications of molecular genetic surveys (first on proteins, later on the DNA directly; see Appendix 1, which gives a brief survey of the molecular techniques used to measure genetic variation). These new molecular techniques allowed genes to be defined biochemically and irrespective of whether or not they had allelic variation. The initial studies (Harris 1966; Johnson et al. 1966; Lewontin and Hubby 1966), using techniques that could only detect mutations causing amino acid changes in protein-coding loci (and only a subset of all amino acid changes at that), revealed that about a third of all protein-coding loci were polymorphic (i.e., a locus with two or more alleles such that the most common allele has a frequency of less than 0.95 in the gene pool) in a variety of species. As our genetic survey techniques acquired greater resolution (Appendix 1), this figure has only gone up.

These genetic surveys have made it clear that many species, including our own, have literally astronomically large amounts of genetic variation. The chapters in Part I of this book will examine how premises 1 and 2 combine to explain great complexity at the population level in terms of the amount of genetic variation and its distribution in individuals, within demes among demes, and over space and time. Because it is now clear that many species have vast amounts of genetic variation, the field of population genetics has become less concerned with the amount of genetic variation and more concerned with its phenotypic and evolutionary significance. This shift in emphasis leads directly into our third and final premise.

Phenotypes Emerge from Interaction of DNA and Environment

A **phenotype** is a measurable trait of an individual (or as we will see later, it can be generalized to other units of biological organization). In Morgan's day, genes could only be identified through their phenotypic effects. The gene was often named for its phenotypic

effect in a highly inbred laboratory strain maintained under controlled environmental conditions. This method of identifying genes led to a simple-minded equation of genes with phenotypes that still plagues us today. Almost daily, one reads about "the gene for coronary artery disease," "the gene for thrill seeking," and so on. Equating genes with phenotypes is reinforced by metaphors appearing in many textbooks and science museums to the effect that DNA is the "blueprint" of life. However, DNA is not a blueprint for anything; that is not how genetic information is encoded or processed. For example, the human brain contains about 10^{11} neurons and 10^{15} neuronal connections (Coveney and Highfield 1995). Does the DNA provide a blueprint for these 10^{15} connections? The answer is an obvious "no." There are only about three billion base pairs in the human genome. Even if every base pair coded for a bit of information, there is insufficient information storage capacity in the human genome by several orders of magnitude to provide a blueprint for the neuronal connections of the human brain. DNA does not provide phenotypic blueprints; instead the information encoded in DNA controls dynamic processes (such as axonal growth patterns and signal responses) that always occur in an environmental context. There is no doubt that environmental influences have an impact on the number and pattern of neuronal connections that develop in mammalian brains in general. It is this interaction of genetic information with environmental variables through developmental processes that yield phenotypes (such as the precise pattern of neuronal connections of a person's brain). Genes should never be equated to phenotypes. Phenotypes emerge from genetically influenced dynamic processes whose outcome depends upon environmental context.

In this book, phenotypes are always regarded as arising from an interaction of genotype with environment. The marine worm *Bonellia* (Figure 1.2) provides an example of this interaction (Gilbert 2000). The free-swimming larval forms of these worms are sexually

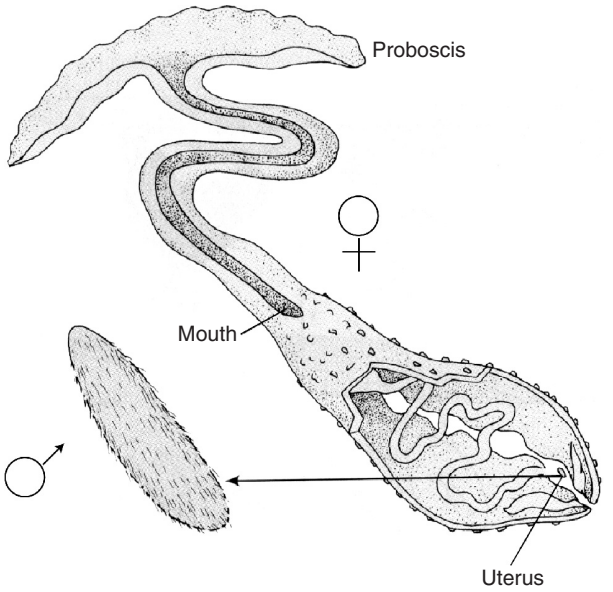

Figure 1.2. Sexes in *Bonellia*. The female has a walnut-sized body that is usually buried in the mud with a protruding proboscis. The male is a ciliated microorganism that lives inside the female. Adapted from Fig. 3.18 from *Genetics*, 3rd Edition, by Peter J. Russell. Copyright © 1992 by Peter J. Russell. Reprinted by permission of Pearson Education, Inc.

undifferentiated. If a larva settles alone on the normal mud substrate, it becomes a female with a long (about 15-cm) tube connecting a proboscis to a more rounded part of the body that contains the uterus. On the other hand, the larva is attracted to females, and if it can find a female, it differentiates into a male that exists as a ciliated microparasite inside the female. The body forms are so different they were initially thought to be totally different creatures. Hence, the same genotype, depending upon environmental context, can yield two drastically different body types. The interaction between genotype and environment in producing phenotype is critical for understanding the evolutionary significance of genetic variability, so the chapters in Part II will be devoted to an exploration of the premise that phenotypes emerge from a genotype-by-environment interaction.

As a prelude to why the interaction of genotype and environment is so critical to evolution, consider the following phenotypes that an organism can display:

- Being alive versus being dead: the phenotype of **viability** (the ability of the individual to survive in the environment)
- Given being alive, having mated versus not having mated; the phenotype of **mating success** (the ability of a living individual to find a mate in the environment)
- Given being alive and mated, the number of offspring produced; the phenotype of **fertility** or **fecundity** (the number of offspring the mated, living individual can produce in the environment)

The three phenotypes given above play an important role in microevolutionary theory because collectively these phenotypes determine the chances of an individual passing on its DNA in the context of the environment. The collective phenotype produced by combining these three components required for passing on DNA is called **reproductive fitness**. Fitness will be discussed in detail in Part III. Reproductive fitness turns premise 1 (DNA can replicate) into reality. DNA is not truly self-replicating. DNA can only replicate in the context of an individual surviving in an environment, mating in that environment, and producing offspring in that environment. Hence, the phenotype of reproductive fitness unites premise 3 (phenotypes are gene-by-environment interactions) with premise 1. This unification of premises implies that the probability of DNA replication is determined by how the genotype interacts with the environment. In a population of genetically diverse individuals (arising from premise 2 that DNA can mutate and recombine), it is possible that some genotypes will interact with the environment to produce more or fewer acts of DNA replication than other genotypes. Hence, the environment influences the relative chances for various genotypes of replicating their DNA. As we will see in Part III, this influence of the environment (premise 3) upon DNA replication (premise 1) in genetically variable populations (premise 2) is the basis for natural selection and one of the major emergent features of microevolution: **adaptation** to the environment, which refers to attributes and traits displayed by organisms that aid them in living and reproducing in specific environments. Adaptation is one of the more dramatic features of evolution, and indeed it was the main focus of the theories of Darwin and Wallace. Adaptation can only be understood in terms of a three-way interaction among all of the central premises of population genetics.

This book uses these three premises in a progressive fashion: Part I utilizes premises 1 and 2, which are molecular in focus, to explain the amount and pattern of genetic variation under the assumption that the variation has no phenotypic significance. Part II focuses upon premise 3 and considers what happens when genetic variation does influence phenotype. Finally, Part III considers the emergent evolutionary properties that arise from the

interactions of all three premises and specifically focuses upon adaptation through natural selection. In this manner, we hope to achieve a thorough and integrated theory of microevolutionary processes.

METHODOLOGICAL APPROACHES IN POPULATION GENETICS

Evolutionary processes have produced an immense array of biological diversity on this planet, with species displaying complex and intricate adaptations to their environments. Understanding this diversity and complexity, its origins, and its implications ranging from the molecular through ecological levels is a daunting challenge. To meet this challenge, the study of population genetics requires an appreciation of a broad range of scientific approaches. We will make use of four approaches in this book:

- Reductionism
- Holism
- Comparative analysis
- Monitoring of natural populations

Reductionism

At one end of the above range of methodologies is the reductionist approach. Reductionism seeks to break down phenomena from a complex whole into simpler, more workable parts to find underlying rules, laws, and explanations. The reductionist approach is based upon the assumption that many complex features of a system can be explained in terms of a few components or rules contained within the system itself; that is, the explanation for the observed complexity lies within the *content* of the system. In this manner, simplicity (the parts contained within the system) generates complexity (the attributes of the whole system). Reductionism seeks necessary and sufficient explanations for the phenomenon under study. Such content-oriented explanations based upon reductionism are said to be proximate causes for the phenomenon of interest.

For example, why do people die? A reductionist approach would look at each instance of death and attempt to describe why that particular person died at that particular time in terms of the status of that individual's body at the time of death. One would get different answers for different individuals, and one would not need to look beyond the health status of a particular individual to obtain the proximate answer. Death is explained exclusively in terms of the content of the individual's body and nothing external to the body is considered. Taking such a reductionist approach, the three leading proximate causes of death in the year 2000 in the United States are (1) heart disease (29.6% of all deaths that year), (2) cancer (23%), and (3) cerebrovascular disease (7%) (Mokdad et al. 2004).

Much of population genetic theory and practice are reductionist in approach. One of the primary tools for implementing the reductionist approach is the controlled experiment in which all potential variables save one are ideally fixed, thereby allowing strong inference about how the single remaining variable factor causes effects of interest in the system under study. The controlled experiment fixes the context to allow inference about the content of a system varying with respect to a single factor. The experimental approach has been widely applied in population genetics and has proven to be a powerful tool in elucidating causal

factors in microevolution. Note, however, that the strong inferences made possible by this approach are limited by the fixed contexts of the experiment, so generalizations outside of that context need to be made with great caution. Moreover, potential interactions with variables that have been experimentally fixed lie outside the domain of inference of the experimental approach. Indeed, in the ideal controlled experiment in which only a single factor is varying, all interaction effects are eliminated from the domain of inference, so some potentially important biological phenomena are not amenable to inference in a controlled experiment.

The reductionist approach is used in both experimental and theoretical population genetics. In modeling microevolution, the complexity of an evolving population is often simplified by reducing the number of variables and ignoring many biological details. With such simplification, laws and complex evolutionary patterns can be elucidated from a few components or factors that are contained within the population itself. Part I uses a reductionist approach to explain the fates and patterns of genetic diversity observed in populations in terms of simple attributes of the population itself. This reductionist approach yields an explanation of many important microevolutionary phenomena, often confirmed by appropriate controlled experiments. However, reductionism alone is insufficient to understand all of microevolution.

Holism

As a complement to the reductionist approach that simplicity generates complexity, the holistic approach is based upon the assumption that simple patterns exist in nature that emerge when underlying complex systems are placed into a particular *context* (simplicity emerges from complexity). The explanation of these emergent patterns often depends not upon knowing the detailed content of the component systems but rather upon the context in which these components are placed in a higher level interacting whole. These context-dependent explanations that do not depend upon detailed content reveal what is commonly called ultimate causation.

For example, why do people die? A holistic approach would look at multiple variables that define the health context of a population of individuals. One would not be trying to explain why a particular individual died at a particular time, but rather one would be trying to access the importance of context variables as predictors of death at the level of the whole population. Taking such a holistic approach, the three leading ultimate causes of death in the year 2000 in the United States are (1) tobacco consumption (18.1%), (2) being overweight (poor diet and physical inactivity, 16.6%), and (3) alcohol consumption (3.5%) (Mokdad et al. 2004). The ultimate explanation of causes of death does not depend upon the cause of death of any particular individual. The ultimate answers as to why people die also depend *not* upon the state of their bodies at the time of death (content) but rather upon the environmental context (tobacco, diet, physical activity, alcohol) into which their bodies have been placed.

It is critical to note that reductionist and holistic approaches are complementary, not antagonistic. Both approaches provide answers that are meaningful, albeit at different biological levels. A practicing physician would be most concerned with the particular health status of his or her patients. Such a physician would be prescribing specific treatments for specific individuals based on studies and knowledge of proximate causation. However, a public health official would focus more on ultimate causation and would try to augment the health of the U.S. population by encouraging less tobacco use and reducing the number of overweight people. Both answers to why people die are valid and both answers can be

used in making health-related decisions. The reductionist and holistic answers each lead to insights and details that are not addressed by the other.

Moreover, reductionist and holistic approaches can converge. A controlled experiment can allow two or more factors to vary, not just one, and can be designed to look at the interactions of the variables. This allows one to study the effect of one variable in the context of another variable. Similarly, a holistic study can be designed that controls (fixes) some variables, resulting in ultimate answers that focus on the content defined by the remaining variables. For example, one can do "case–control" studies by assembling two groups of people, say one group of smokers and one group of nonsmokers, who are matched on several other variables (age, gender, etc.). Such studies have revealed that smoking increases the risk of individuals developing heart disease, cancer, and cerebrovascular disease, thereby forging a link between the studies on ultimate and proximate causations of death in the U.S. population. In this manner, the gap between reductionism and holism and between proximate and ultimate causation can often be narrowed.

All too often, reductionism and holism are presented as alternative, antagonistic approaches in biology. This legacy is particularly true for studies on the inheritance of traits, which has often been phrased as a debate between nature (content) and nurture (context). As discussed earlier in this chapter, this is a false dichotomy. Premise 3 tells us that traits emerge from the *interaction* of genotypes with environments, and modern studies on trait variation often seek to examine both content (the genes affecting trait variation) and context (the environments in which the genes are expressed). As soon as we deal with the phenotypic significance of genetic variation in Part II, an exclusively molecular, reductionist focus is no longer appropriate. Rather we must take an organismal, holistic focus in the context of an environment.

Of the traits that emerge from the interaction of genotypes with environment are those traits related to the ability of an individual to reproduce and pass on genes to the next generation. As already discussed in this chapter and in detail in Part III, the evolutionary mechanism of natural selection emerges from the interaction of genotypes with environments. Many explanations of ultimate causation in evolutionary biology depend upon natural selection. Again and again, the traits expressed by particular individuals or in particular populations or species are explained in the ultimate sense in terms of arguments of how particular environmental contexts result in natural selection favoring the trait. Population genetics deals in part with the mechanism of natural selection (Part III), and hence population genetics is an essential component of any explanation of ultimate causation based upon evolutionary change induced by natural selection. However, the population genetic approach to mechanisms such as natural selection explicitly uses both reductionism and holism simultaneously. For example, in population genetics natural selection is discussed in terms of the specific genes *contained* within the organisms being selected and the mapping of these genes to phenotype in the context of an environment, with the evolutionary response modulated by the other evolutionary forces contained within the population as discussed in Part I. Such an integrated reductionist/holistic approach will be the emphasis in Parts II and III.

Comparative Analysis

An evolutionary process occurs over time; therefore evolving populations (and the genes contained within those populations) have a history. The comparative approach to biological science makes active use of this history. This is a scientific method used extensively in biology, mostly at the species level and above. Traditionally, an evolutionary tree is

constructed for a group of species. Then other data about these organisms (anatomy, developmental pathways, behavior, etc.) are overlaid upon the evolutionary tree. In this manner, it is possible to infer how many evolutionary transitions occurred in characters of interest, the locations of transitions within the evolutionary tree, and patterns of evolutionary associations among characters. Contrasts between those organisms on either side of a transitional branch are those that are most informative about the character of interest because the sharing of evolutionary history for all other traits is maximized by this contrast. A comparative contrast bears some similarity to a controlled experiment in reductionist empirical science because the contrast is chosen to minimize confounding factors.

For example, Darwin's finches comprise a group of 14 species of songbirds living on the Galápagos Islands and Cocos Island off the coast of Equador that were collected by Charles Darwin and other members of the *Beagle* expedition in 1835. These 14 species have drawn the attention of many evolutionary biologists because of the remarkable diversity in the shape and size of their beaks, which range from sharp and pointed to broad and deep (Figure 1.3). Why do these 14 species show such remarkable diversity in beak shape and size? Both the proximate and ultimate answers to this question have been studied using the comparative method. Petren et al.(1999) estimated an evolutionary tree of these finches from molecular genetic differences, with the resulting tree shown in Figure 1.3. Abzhanov et al. (2004) compared beak development in the six species of the genus *Geospiza* from this evolutionary tree and also compared the expression patterns of a variety of growth factors that are known to influence avian craniofacial development. By overlaying these data upon the molecular genetic tree, they produced evolutionary contrasts that separated out the effects of beak size and beak shape. Most of the growth factors they examined showed no significant pattern of change on this evolutionary tree. The expression patterns of bone morphogenetic proteins 2 and 7, coded for by the *Bmp2* and *Bmp7* genes, respectively, correlated with beak size but not with beak shape. The expression patterns of bone morphogenetic factor 4, coded for by the *Bmp4* gene, strongly correlated with beak shape changes on this evolutionary tree. Because the comparative study implicated *Bmp4* expression as being an important proximate cause of beak shape diversity, Abzhanov et al. (2004) next performed controlled experiments to test this hypothesis within a reductionist framework. They attached the chicken *Bmp4* gene to a viral vector and infected developing cells with this virus to alter the expression of the *Bmp4* gene. In this manner, they were able to alter the beak shape of chick embryos in a manner that mimicked the types of changes observed in the evolution of Darwin's finches.

This work on *Bmp4* expression does not, however, provide the ultimate answer as to why Darwin's finches show much diversity in beak size and shape. The comparative approach can also be used to address the ultimate question of what environmental factors, if any, caused this beak diversity and underlying patterns of *Bmp4* expression to have evolved through natural selection. Perhaps this beak diversity evolved on the South American mainland, and the Galápagos Islands were simply colonized by finches with preexisting beak diversity. In this case, the ultimate answer would lie in evolution in the mainland and have little to do with the context of being on the Galápagos Islands. Alternatively, if all 14 species evolved on the Galápagos Islands, then the ultimate answer would lie specifically in the context of the Galápagos Islands. The evolutionary tree in Figure 1.3 shows that all 14 species evolved on the Galápagos Islands, so the ultimate answer should lie in the environments found on these islands. This shows that just having an evolutionary tree allows some hypothesis about ultimate causation to be tested directly. The comparative analysis clearly indicates that the

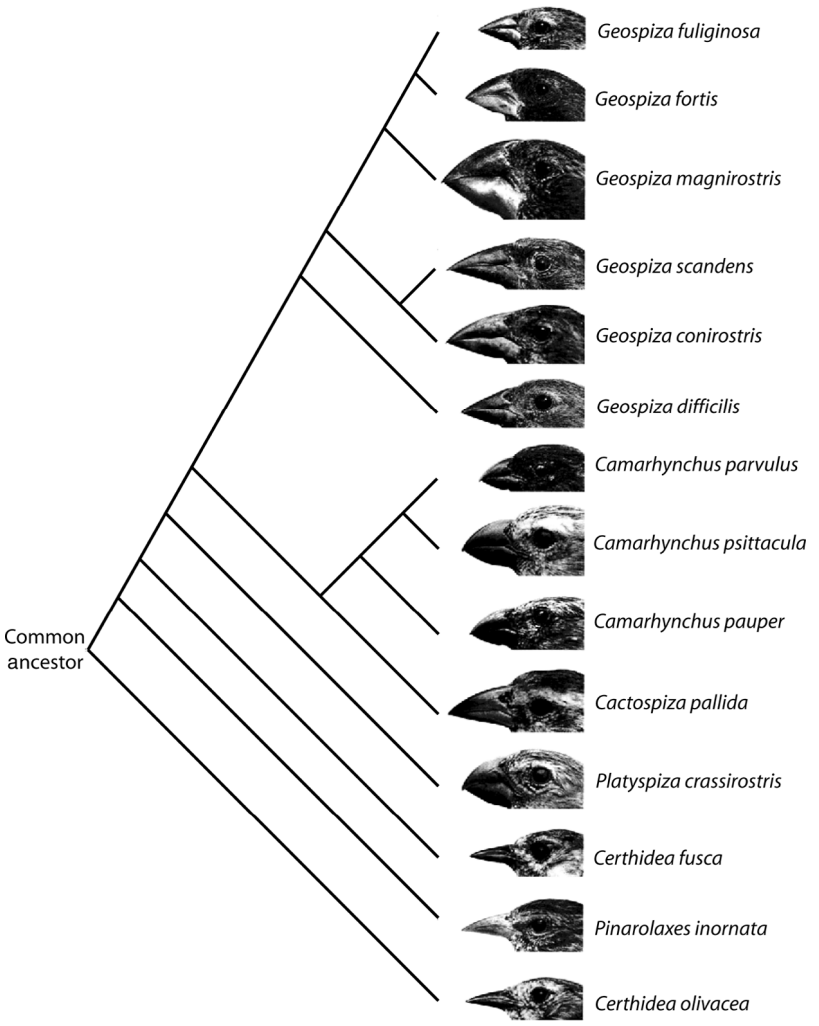

Figure 1.3. Evolutionary tree of 14 species of Darwin's finches estimated from molecular genetic data. Modified from Fig. 3 in Petren et al. (1999). Copyright ©1999 by the Royal Society of London.

ultimate explanation lies in the environments found on the Galápagos Islands and not on the mainland.

Because beaks are used to procure and process food, diet is a logical environmental factor for studies on how natural selection may have shaped beak diversity in these finches. Fieldwork has revealed much about the ecology of Darwin's finches (Grant 1986), including their diets. Different species eat items of different sizes, an example of which is shown in Figure 1.4. This dietary data can also be overlaid upon the evolutionary tree of the finches, and it reveals a strong correlation in shifts of diet with transitions in beak size and shape. Note that this comparative analysis reveals a strong association between content (the beak size and shape of individual species) and context (the dietary environment). Such a

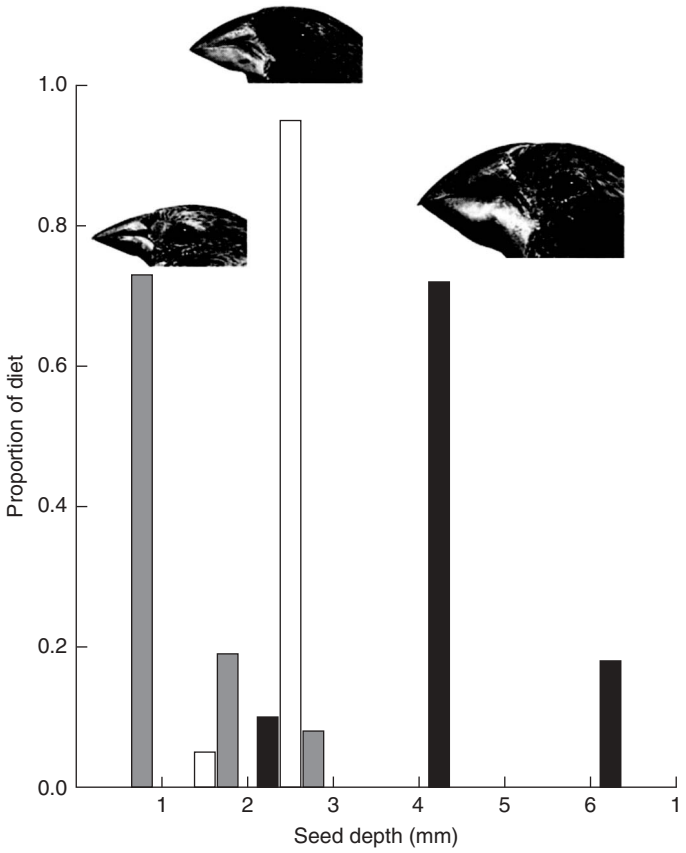

Figure 1.4. Proportions of various seed sizes in diet of three of Darwin's finches: *Geospiza magnirostris* (solid bars), *Geospiza fortis* (open bars), and *Geospiza fuliginosa* (gray bars). Redrawn with permission from Fig. 35 in P. R. Grant, *Ecology and Evolution of Darwin's Finches* (1986). Copyright ©1986 by Princeton University Press.

content–context association in evolutionary history suggests the hypothesis that the beak diversity is being shaped by natural selection as adaptations for different diets.

These studies on Darwin's finches reveal that the comparative approach can be used to test and formulate hypotheses of both proximate and ultimate causation. One of the more exciting developments in population genetics during the last part of the twentieth century was the development of molecular techniques that have allowed the application of comparative approaches *within* species. As illustrated above, it is now possible to trace the evolutionary history of species through molecular genetic studies. However, this evolutionary history can often be inferred for the genetic variation found within a species as well. In this manner, population genetic studies on genetic variation within a species can now include the evolutionary history of that genetic variation. This opens the door to comparative approaches within species. Such intraspecific comparative approaches are used throughout this book, and they represent a particularly powerful way of uniting reductionism and holism within population genetics.

Monitoring Natural Populations

Many hypotheses in population genetics can be tested by monitoring natural populations. One of the simplest types of monitoring is a one-time sample of individuals of unknown relationship coupled with some sort of genetic survey (using one or more of the techniques described in Appendix 1). Such simple genetic surveys allow one to estimate and test most of the evolutionary forces described in Part I. Just as genetic surveys of present-day species can allow an evolutionary tree of those species to be estimated (e.g., Figure 1.3), so can a genetic survey of present-day genes and/or populations allow an evolutionary history of those genes and/or populations to be estimated. Moreover, the genetic survey data can be overlaid with phenotypic data to test hypothesis about how genetic variation influences phenotypic variation, as will be shown in Part II. Finally, Part III shows that many tests for the presence or past operation of natural selection are possible from such genetic survey data.

The monitoring of natural populations can be extended beyond a simple one-time survey of genetic variation of individuals of unknown relationship. For example, one can sample families (parents and offspring) instead of individuals or follow a population longitudinally through time to obtain multigeneration data. Such designs allow more hypotheses to be tested. For example, Boag (1983) sampled parents and offspring of the Darwin finch *Geospiza fortis* and plotted the beak depth of the offspring against the average beak depth of their two parents (Figure 1.5). As will be shown in Part II, such data can be used to

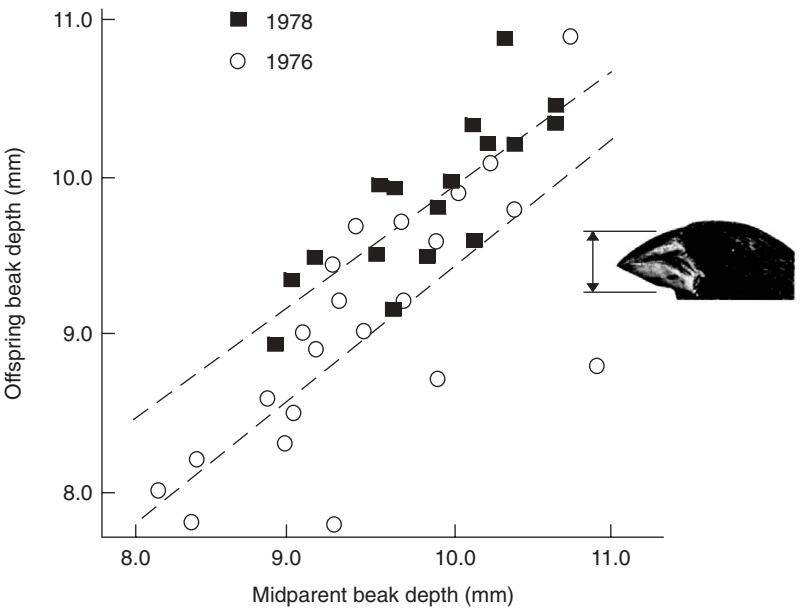

Figure 1.5. Relationship between beak depth of offspring and average beak depth of their parents (midparent beak depth) in medium ground finch, *G. fortis*, as measured in two years, 1976 and 1978. The lines show a fitted least-squares regression to these data (Appendix 2). As will be explained in Chapter 9, the nearly identical, positive slopes of these lines indicate that genetic variation in these populations contribute in a major way to variation in beak depth. From Fig. 1 in Boag (1983). Copyright ©1983 by The Society for the Study of Evolution.

estimate the contribution of genetic variation to variation in the trait of beak depth even in the absence of a molecular genetic survey. In this case, the plots shown in Figure 1.5 reveal that the intraspecific variation observed in beak depth in *G. fortis* is strongly influenced by genetic variation within this population.

Population genetics is concerned with the fate of genes over space and time within a species, and this fate can be observed or estimated by monitoring populations over space and time. Such monitoring over space and time also allows population geneticists to make use of natural experiments. For example, natural selection arises out of how individuals interact with their environment, but environments themselves often change over space and time. Although not a controlled experiment in the strict reductionist sense, spatial and temporal environmental contrasts can sometimes provide a similar inference structure. To see how, consider again Darwin's finches. The comparative method implied that the variation in beak size and shape reflected adaptations to dietary differences. However, this answer of ultimate causation raises yet other questions about ultimate causation: Why did some or all of the current species evolve a different diet from that of the common ancestral finch and why do the current species display such a variety of diets? These questions of ultimate causation can be addressed through the use of natural experiments involving environmental contrasts in time and space. For example, in 1977 the Galápagos Islands suffered a severe drought. By monitoring both the finch populations and the environment in which they live, it was discovered that this drought had a major impact on both the abundance of the seeds eaten by these finches and the characteristics of the seeds. For example, there was a dramatic shift from small and soft seeds to large and hard seeds during the drought for the seeds eaten by the medium ground finch, *G. fortis* (Figure 1.6). The inference from the comparative method that beak size and shape are adaptive to diet leads to the prediction that this drought-induced shift in diet would result in natural selection on the beaks in *G. fortis*. This prediction is testable by monitoring the population before and after the drought. There

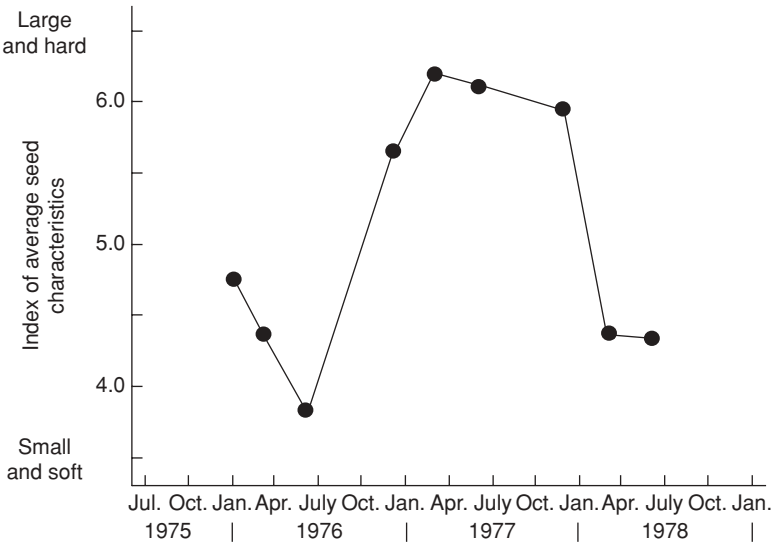

Figure 1.6. Characteristics of average seed available as food to medium ground finches (*G. fortis*) before, during, and after 1977 drought. Reprinted from Fig. 1 in P. T. Boag and P. R. Grant, *Science* 214: 82–85(1981). Copyright ©1981 by the AAAS.

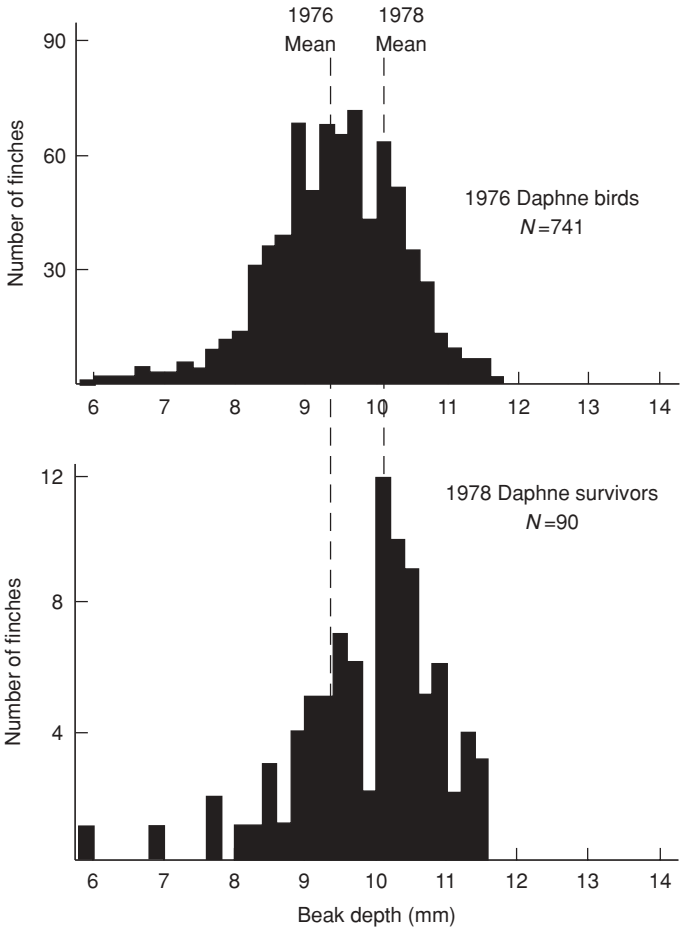

Figure 1.7. Frequency distributions of beak depth in *G. fortis* on island of Daphne Major before (1976) and after (1978) a drought. Dashed lines indicate the mean beak depths in 1976 and 1978. Redrawn with permission from Fig. 59 in P. R. Grant, *Ecology and Evolution of Darwin's Finches* (1986). Copyright © 1986 by Princeton University Press.

was a significant shift upward in beak depth in the survivors of the drought relative to the predrought population (Figure 1.7), a shift consistent with the hypothesis that increased beak depth is an adaptation to the larger and harder seeds that were available during the drought. Given that variation in beak depth is strongly influenced by genetic variation in this population (Figure 1.5), another prediction is that natural selection operated on this population to cause evolution in this population in response to the drought. This prediction can also be tested by looking at the beak depths of the finches hatched in the years before and after the drought, and indeed the predicted genetic shift is observed (Figure 1.8).

Subsequent environmental changes confirmed that changes in seed availability induce selection on beak shape and size (Grant and Grant 1993, 2002). Even though the subsequent environmental shifts were different from those induced by a drought, this environmental heterogeneity over time did replicate the testable prediction that beak shape and size are subject to natural selection due to interactions with the available seed environment. These

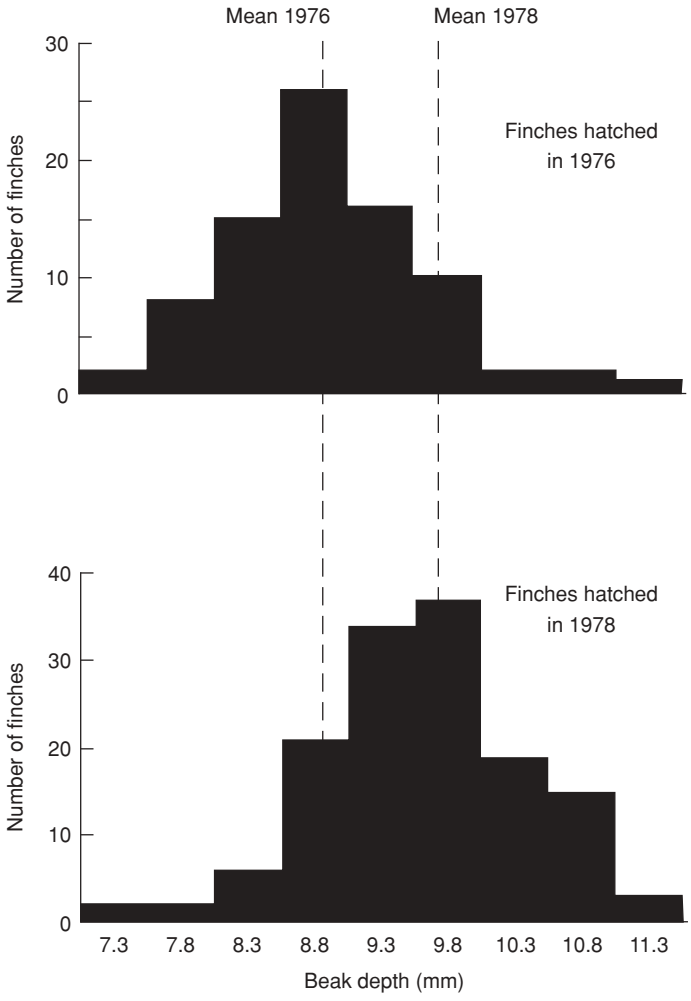

Figure 1.8. Beak depth in *G. fortis* hatched year before drought (1976) versus year after drought (1978). Dashed lines indicate the mean beak depth for the finches born before and after the drought.

natural experiments from monitoring populations reinforce the inference made from the comparative method that beak size and shape are adaptations to diet. Moreover, these temporal natural experiments suggest that beak size and shape would not remain static once an ancestral finch colonized these islands but rather would evolve because the seed environment is subject to change over time. Moreover, the seed environment varies from island to island, so this selective hypothesis could also explain some of the diversity of beak size and shape between finch species that primarily live on different islands.

These studies on Darwin's finches illustrate that the monitoring of natural populations can be a powerful method of inference in population genetics. Note that studies on Darwin's finches have utilized reductionist controlled experiments, reductionist comparative studies, holistic comparative studies, and monitoring of natural populations. The best studies in population genetics tend to integrate multiple methods of inference that are complementary and reinforcing to one another.

I

POPULATION STRUCTURE AND HISTORY

2

MODELING EVOLUTION AND THE HARDY–WEINBERG LAW

Throughout this book we will construct models of reproducing populations to investigate how various factors can cause evolutionary changes. In this chapter, we will construct some simple models of an isolated local population. These models use a reductionistic approach to eliminate many possible features in order to focus our inference upon one or a few potential microevolutionary factors. The models will also provide insights that have been historically important to the acceptance of the neo-Darwinian theory of evolution at the beginning of the twentieth century and are of increasing importance to the application of genetics to human health and other contemporary problems at the beginning of the twenty-first century.

HOW TO MODEL MICROEVOLUTION

Given our definition that evolution is a change over time in the frequency of alleles or allele combinations in the gene pool, any model of evolution must include at the minimum the passing of genetic material from one generation to the next. Hence, our fundamental time unit will be the transition between two consecutive generations at comparable stages. We can then examine the frequencies of alleles or allelic combinations in the parental versus offspring generation to infer whether or not evolution has occurred. All such transgenerational models of microevolution have to make assumptions about three major mechanisms:

- Mechanisms of producing gametes
- Mechanisms of uniting gametes
- Mechanisms of developing phenotypes.

In order to specify how gametes are produced, we have to specify the genetic architecture. **Genetic architecture** refers to the number of loci and their genomic positions, the number

Population Genetics and Microevolutionary Theory, By Alan R. Templeton

of alleles per locus, the mutation rates, and the mode and rules of inheritance of the genetic elements. For example, the first model we will develop assumes a genetic architecture of a single autosomal locus with two alleles with no mutation. The genetic architecture provides the information needed to specify how gametes are produced. For a single-locus, two-allele autosomal model with no mutation, we need only to use Mendel's first law of inheritance (the law of equal segregation of the two alleles in an individual heterozygous at an autosomal locus) to specify how genotypes produce gametes. Other single-locus genetic architectures can display different modes of inheritance, including X-linked loci (with a haplo–diploid, sex-linked mode of inheritance), Y chromosomal loci (with a haploid, unisexual paternal mode of inheritance in humans), or mitochondrial DNA (with a haploid, maternal mode of inheritance in humans). We can also examine genetic architectures that depend upon more than one locus, in which case mixed modes of inheritance are possible and in which Mendel's second law (independent assortment) and/or recombination frequencies of linked loci may enter into the rules by which gametes are produced. We can even have deviations from the standard rules of inheritance. For example, we may specify that a locus is subject to deviations from Mendel's first law of 50–50 segregation in the production of gametes from heterozygotes. In a multilocus model we may specify that unequal crossing over can occur, thereby producing variation in the number of genes transmitted to the gametes. The assumptions about genetic architecture that we make obviously limit the types of evolutionary processes that we can model. Hence, the specification of genetic architecture is a critical first step in any model of microevolution.

Because our focus is upon sexually reproducing diploid organisms, the transition from one generation to the next involves not only the production of gametes but also the pairing of gametes to form new diploid zygotes. Hence, we need to specify the mechanisms or rules by which gametes are paired together in the reproducing population. These mechanisms of uniting gametes are called **population structure**. **Population structure** includes the following:

- System of mating of the population
- Size of the population
- Presence, amount, and pattern of genetic exchange with other populations
- Age structure of the individuals within the population

All of these factors can have an impact on which gametes are likely to be paired and transmitted to the next generation through newly formed zygotes. As with genetic architecture, we can make assumptions about population structure that vary from the simple to the complex, depending upon the types of phenomena we wish to examine for evolutionary impact. The system of mating can be simply a random pairing of individuals or can be influenced by degrees of biological relatedness or other factors. We can choose to ignore the impact of population size by assuming size to be infinite or we can examine small populations in which the population size has a major impact on the probability of two gametes being united in a zygote. We can model a single deme in which all uniting gametes come from that deme or we can allow gametes from outside the deme to enter at some specified rate or probability, which in turn could be a function of geographical distance, ecological barriers, and so on. We can assume discrete generations in which all individuals are born at the same time and then reproduce at the same time followed by complete reproductive senescence or death. Alternatively, we can assume that individuals can reproduce at many times throughout their life and can mate with individuals of different ages and offspring can coexist with their parents. Until we specify these parameters of population structure,

we cannot model microevolution because the uniting of gametes is a necessary step in the transmission of genes from one generation to the next in sexually reproducing organisms.

In most species, the zygote that results from uniting gametes is not capable of immediate reproduction but rather must grow, develop, survive, and mature reproductively. All of this takes place in an environment or suite of environments. From premise 3 in Chapter 1 (phenotypes are gene-by-environment interactions), we know that actual DNA replication depends upon the phenotypes of the individuals bearing the DNA. Hence, we also need to specify **phenotypic development**, that is, the mechanisms that describe how zygotes acquire phenotypes in the context of the environment. Assumptions can range from the simple (the genetic architecture has no impact on phenotype under any of the environments encountered by individuals in the population) to the complex (phenotypes are dynamic entities constantly changing as the external environment changes and/or as the individual ages with changing patterns of epistasis and pleiotropy throughout).

All models of microevolution must make assumptions about the mechanisms of producing gametes, uniting gametes, and developing phenotypes. Without such assumptions, it is impossible to specify the genetic transition from one generation to the next. Quite often, models are presented that do not explicitly state the assumptions being made about all three mechanisms. This does not mean that assumptions are not being made; rather, they are being made in an implicit fashion. Throughout this book an effort is made to state explicitly the assumptions being made about all three of these critical components of transferring DNA from one generation to the next in a reproducing population. We will do this now for our first and simplest model of evolution, commonly called the Hardy–Weinberg model.

HARDY–WEINBERG MODEL

One of the simplest models of population genetics is the Hardy–Weinberg model, named after two individuals who independently developed this model in 1908 (Hardy 1908; Weinberg 1908). Although this model makes several simplifying assumptions that are unrealistic, it has still proven to be useful in describing many population genetic attributes and will serve as a useful base model in the development of more realistic models of microevolution. Hardy was an English mathematician, and his development of the model is mathematically simpler but yields less biological insight than the more detailed model of Weinberg, a German physician. Both derivations will be presented here because each has advantages over the other for particular problems that will be addressed later in this book.

Both derivations start with a common set of assumptions, as summarized in Table 2.1. We now discuss each of the assumptions given in that table. Concerning the *mechanisms of producing gametes*, both men assumed a single autosomal locus with two alleles and with no mutation. Meiosis was assumed to be completely normal and regular, so that Mendel's first law of equal segregation could predict the gametes produced by any genotype. There are also no maternal or paternal effects of any sort, so it makes no difference which parent contributes a gamete bearing a specific allele.

Concerning the *mechanisms of uniting gametes*, both men assumed a single population that has no genetic contact with any other populations; that is, an isolated population. Within this closed population, Hardy assumed the individuals are monoecious (each individual is both a male and a female) and self-compatible; Weinberg allowed the sexes to be separate but assumed that the sex of the individual has no impact on any aspect of inheritance or genetic architecture. The system of mating in both derivations is known as **random mating**

Table 2.1. Assumptions of Hardy–Weinberg Model

Mechanisms of producing gametes (genetic architecture)	One autosomal locus, two alleles, no mutation, Mendel's first law
Mechanisms of uniting gametes (population structure)	
System of mating	Random
Size of population	Infinite
Genetic exchange	None (one isolated population)
Age structure	None (discrete generations)
Mechanisms of developing phenotypes	All genotypes have identical phenotypes with respect to their ability for replicating their DNA

and means that the probability of two genotypes being mates is simply the product of the frequencies of the two genotypes in the population. Note that random mating is defined solely in terms of the genotypes at the locus of interest; there is no implication in this assumption that mating is random for any other locus or set of loci or for any phenotypes not associated with the locus of interest. For example, humans do not mate at random for a number of phenotypes (gender, skin color, height, birthplace, etc.), but as long as the genetic variation at the locus of interest has no impact on any of these phenotypes, the assumption of random mating can still hold. Hence, random mating is an assumption that is specific to the genetic architecture of interest and that does not necessarily generalize to other genetic systems found in the same organisms.

Concerning the other aspects of population structure, both derivations make the assumption that the population is of infinite size, thereby eliminating any possible effects of finite population size upon the probability of uniting gametes. Both men ignored the effects of age structure by assuming discrete, nonoverlapping generations. Finally, concerning the mechanisms of developing phenotypes, nothing was explicitly assumed, but implicitly both derivations require that under the range of environments in which the individuals of the population are living and reproducing there is no phenotypic variation for viability, mating success, and fertility. In terms of their ability to replicate DNA, all genotypes have identical phenotypes. This means that all genotypes have the same reproductive fitness, so there is no natural selection in this model.

To examine the population genetic implications of these assumptions upon a reproducing population, we need to go through a complete generation transition. In both derivations, we will start with a population of reproductively mature adults. The essence of this model (and many others in population genetics) is to follow the fate of genes from this population of adults through producing gametes, mating to unite gametes (zygote production), and then zygotic development to the adults of the next generation. We will then examine the gene pools associated with these two generations of adults to see if any evolution has occurred.

Because we are dealing with a single autosomal locus with two alleles (say A and a) and no additional mutation, adult individuals are of three possible genotypes: AA, Aa, and aa. We will characterize the adult population by their genotypes and the frequencies of these genotypes in the total population (see Figure 2.1). Let these three genotype frequencies be G_{AA}, G_{Aa}, and G_{aa}, where the subscript indicates the genotype associated with each

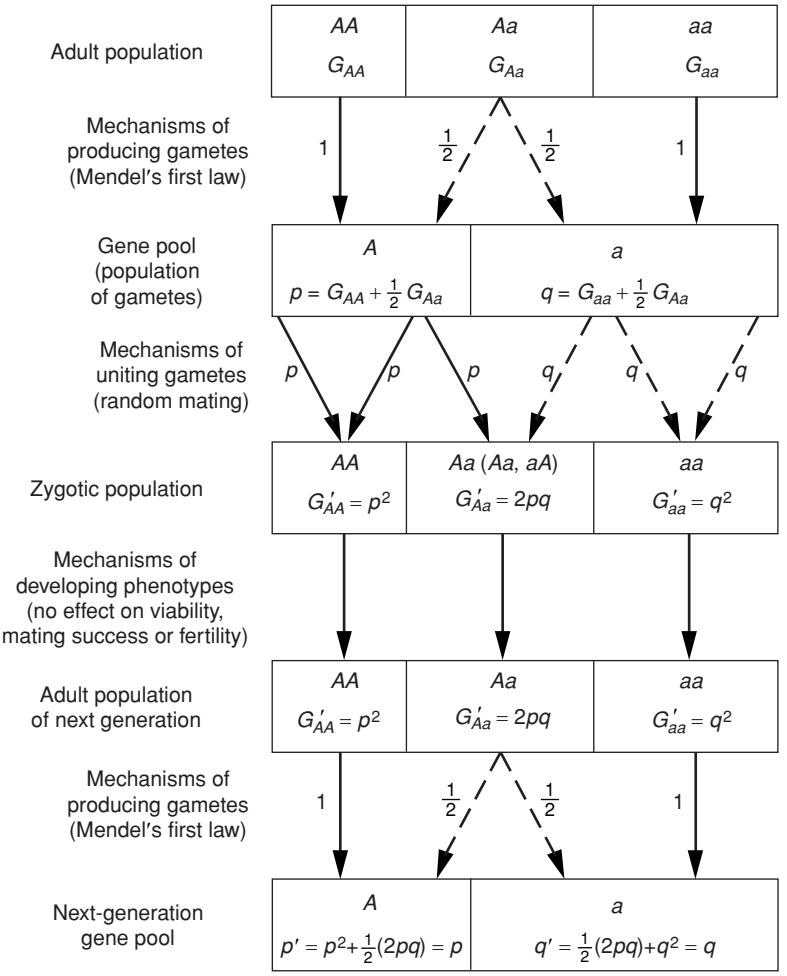

Figure 2.1. Derivation of Hardy–Weinberg law for single autosomal locus with two alleles, *A* and *a*. In going from adults to gametes, solid lines represent Mendelian transition probabilities for homozygotes, and dashed lines represent Mendelian transition probabilities for heterozygotes. In going from gametes to zygotes, solid lines represent gametes bearing the *A* allele, and dashed lines represent gametes bearing the *a* allele.

frequency. Because these three genotypes represent a mutually exclusive and exhaustive set of possible genotypes, these three genotype frequencies define a probability distribution over the genotypes found in the adult population (see Appendix 2 for a discussion of probability distributions). This means that $G_{AA} + G_{Aa} + G_{aa} = 1$. This probability distribution of *genotype frequencies* represents our fundamental description of the *adult population*.

At this point, the derivations of Hardy and of Weinberg diverge. We will first follow Hardy's and then return to Weinberg's. The population of adult individuals can produce gametes. As discussed in Chapter 1, the population of potential gametes produced by these individuals defines the gene pool (Figure 2.1). Because of our assumptions about genetic architecture and no mutation, all we need is Mendel's first law to predict the frequencies of the various haploid genotypes (gametes) found in the gene pool from the frequencies of the

diploid adult genotypes. Two and only two haploid gametic types are possible: A and a. The frequencies of these gametes (which for a one-locus model are called **allele frequencies**) also define a probability distribution over the gamete types found in the gene pool. This probability distribution of *gamete frequencies* represents our fundamental description of the *gene pool*. We will let p be the frequency of gametes bearing the A allele in the gene pool and q the frequency of gametes bearing the a allele in the gene pool. Because p and q define a probability distribution over the gene pool, $p + q = 1$, or $q = 1 - p$. Hence, we need only one number, say p, to completely characterize the gene pool in this model. A critical question is: Can we predict the allele (gamete) frequencies from the genotype frequencies? Under our assumptions of the mechanisms for producing gametes, the answer is "yes" and all we need to use is Mendel's first law of equal segregation. Under Mendel's law, the probability of an AA genotype producing an A gamete is 1 and the probability of an AA genotype producing an a gamete is 0. Similarly, the probability of an aa genotype producing an A gamete is 0 and the probability of an aa genotype producing an a gamete is 1 under standard Mendelian inheritance. Finally, Mendel's first law predicts that the probability of an Aa genotype producing an A gamete is $\frac{1}{2}$ and the probability of an Aa genotype producing an a gamete is $\frac{1}{2}$. These Mendelian probabilities are transition probabilities that describe how one goes from adult genotypes to gamete types. Hence, the transition from the adult population to the gene pool is determined completely by these transmission probabilities (our mathematical descriptor of the mechanisms of producing gametes). As can be seen from Figure 2.1, these transition probabilities from diploidy to haploidy allow us to predict the gene pool state completely from the adult population genotype state. In particular, all we have to do is multiply each transmission probability by the frequency of the genotype with which it is associated and then sum over all genotypes for each gamete type. Thus, $1 \times G_{AA}$ is the frequency of A gametes coming from AA individuals, $\frac{1}{2} \times G_{Aa}$ is the frequency of A gametes coming from Aa individuals, and $0 \times G_{aa} = 0$ is the frequency of A gametes coming from aa individuals. Hence, the total frequency of the A allele in the gene pool is $1 \times G_{AA} + \frac{1}{2} \times G_{Aa} + 0 \times G_{aa} = G_{AA} + \frac{1}{2}G_{Aa} = p$. Similarly, the frequency of the a allele in the gene pool is $0 \times G_{AA} + \frac{1}{2} \times G_{Aa} + 1 \times G_{aa} = G_{aa} + \frac{1}{2}G_{Aa} = q = 1 - (G_{AA} + \frac{1}{2}G_{Aa}) = 1 - p$ (see Figure 2.1). Note that the Mendelian transmission probabilities (the 0's, 1's, and $\frac{1}{2}$'s used above) and the genotype frequencies (the G's) completely determine the allele frequencies in the gene pool. In general, gamete frequencies can always be calculated from genotype frequencies given a knowledge of the mechanisms of producing gametes. Letting g_j be the frequency of gamete type j in the gene pool (either an allele for a single-locus genetic architecture or a multiallelic gamete for a multilocus genetic architecture), the general formula for calculating a gamete frequency is

$$g_j = \sum_{\text{genotypes}} \text{probability (genotype } k \text{ producing gamete } j) \times (\text{frequency of genotype } k) \tag{2.1}$$

where "genotype k" is simply a specific genotype possible under the assumed genetic architecture. The equations previously used to calculate p and q are special cases of equation 2.1 for a single autosomal locus with two alleles. This equation makes it clear that two types of information are needed to calculate gamete frequencies:

- Information about the mechanisms of producing gametes which determine the probability of a specific genotype producing a specific gamete type
- Genotype frequencies of the population of interest

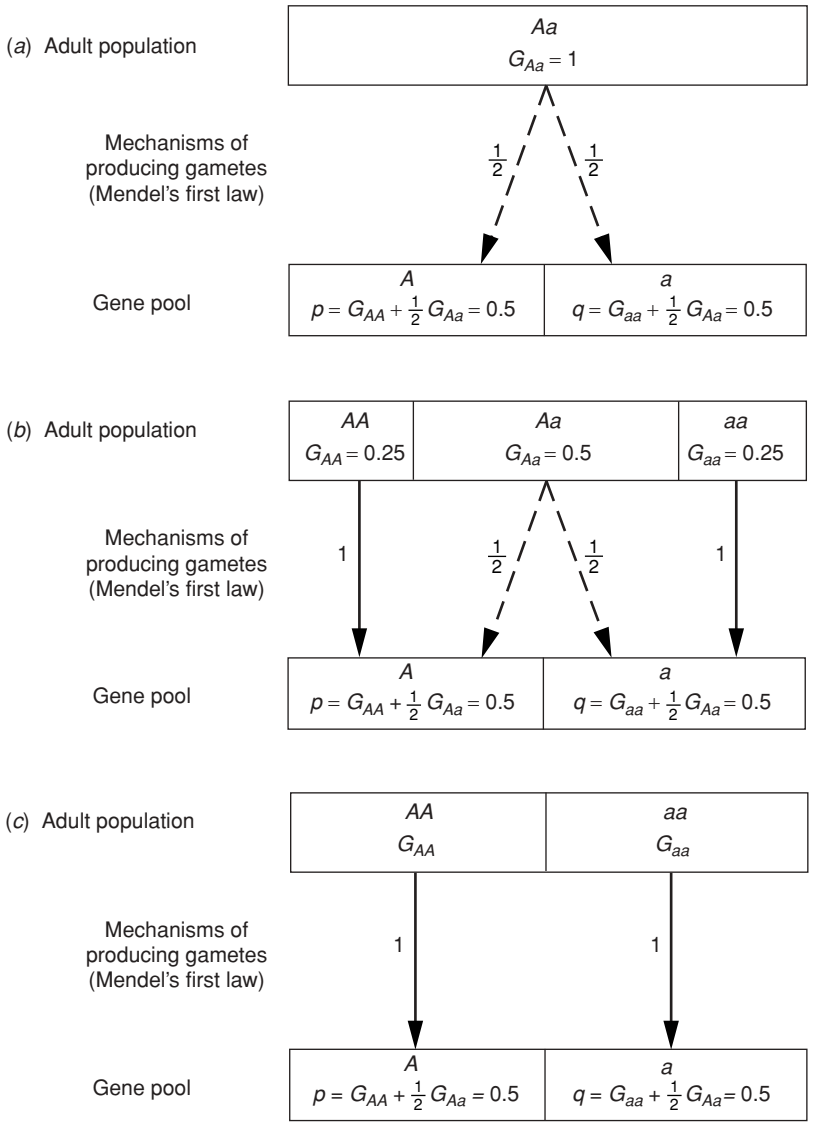

Figure 2.2. Different adult populations sharing a common gene pool.

It is always possible to calculate the gamete frequencies from the genotype frequencies given a knowledge of the mechanisms of producing gametes. Is it also possible to calculate the genotype frequencies from the gamete frequencies given a knowledge of the mechanisms of producing gametes? The answer is "no." To see this, consider a population of adults consisting only of Aa individuals (Figure 2.2a). In this population, $G_{AA} = 0$, $G_{Aa} = 1$, and $G_{aa} = 0$. Hence, $p = G_{AA} + \frac{1}{2} \times G_{Aa} = 0 + \frac{1}{2} \times 1 = 0.5$. Now, consider a population with $G_{AA} = 0.25$, $G_{Aa} = 0.5$, and $G_{aa} = 0.25$ (Figure 2.2b). For this population, $p = G_{AA} + \frac{1}{2} \times G_{Aa} = 0.25 + \frac{1}{2} \times 0.5 = 0.5$. Now consider the population shown in Figure 2.2c, in which $G_{AA} = 0.5$, $G_{Aa} = 0$, and $G_{aa} = 0.5$. In this population, $p = G_{AA} + \frac{1}{2} \times G_{Aa} = 0.5 + \frac{1}{2} \times 0 = 0.5$. Hence, three very different populations of

adults all give rise to identical gene pools! This shows that there is no one-to-one mapping between genotype frequencies and gamete frequencies. Although gamete frequencies can always be calculated from genotype frequencies given a knowledge of the rules of inheritance, genotype frequencies are *not* uniquely determined by gamete frequencies and the rules of inheritance. Obviously, we need additional information to predict genotype frequencies from gamete frequencies. This is where population structure comes in.

Hardy and Weinberg made assumptions about population structure that remove as potential evolutionary factors mutation, genetic contact with other populations, population size, and age structure. All that is left in their simplified model is system of mating. Under Hardy's formulation, random mating means that two gametes are randomly and independently drawn from the gene pool and united to form a zygote. By a random draw, Hardy meant that the probability of a gamete being drawn is the same as its frequency in the gene pool. Hence, if the proportion of the gametes bearing the A allele is p, then the probability of choosing a gamete with an A allele is p. Similarly, the probability of drawing an a gamete is q. Individuals are monoecious in Hardy's model, and every individual contributes equally to both male and female gametes. Hence, although the second gamete drawn from the gene pool must be from the opposite sex of the first, all individuals are still equally likely to be the source of the second gamete. Moreover, Hardy regarded the number of gametes that could be produced by an individual as effectively infinite, so that drawing the first gamete from the gene pool has no effect upon drawing the second. The assumption of random mating also stipulates that this second gamete is drawn independently from the gene pool, which means that the probabilities are identical on the second draw and that the joint probability of both gametes is simply the product of their respective allele frequencies. Table 2.2 shows how these gamete frequencies are multiplied to yield zygotic genotype frequencies. Note, in calculating the frequency of the Aa genotype, there are two ways of creating a heterozygous zygote; the A allele could come from the paternal parent and a from the maternal or vice versa. The Hardy–Weinberg assumptions imply that parental origin of an allele has no

Table 2.2. Multiplication of Allele Frequencies to Yield Zygotic Genotypic Frequencies under Hardy–Weinberg Model of Random Mating

		Male gametes	
	Allele	A	a
	Frequency	p	q

	Allele	Frequency
Female gametes	A	p
	a	q

AA $p \times p = p^2$	Aa $p \times q = pq$
aA $q \times p = qp$	aa $q \times q = q^2$

Summed frequencies in zygotes:

AA: $G'_{AA} = p^2$

Aa: $G'_{Aa} = pq + qp = 2pq$

aa: $G'_{aa} = q^2$

Note: The zygotic genotype frequencies are indicated by G'_k

effect. Hence, the two types of heterozygotes, each with frequency pq, are pooled together into a single Aa class with frequency $2pq$.

As the zygotes develop and mature into adults capable of contributing genes to the next generation, there is no change in their relative frequencies because of the implicit assumption of no phenotypic variation in viability, mating success, or fertility. Hence, Hardy showed that the genotype frequencies of the next generation could be predicted from allele frequencies given knowledge of the system of mating. From Figure 2.1 or Table 2.2, these predicted genotype frequencies are

$$G'_{AA} = p^2 \qquad G'_{Aa} = 2pq \qquad G'_{aa} = q^2$$

This array of genotype frequencies is known as the **Hardy–Weinberg law**.

We did not make any assumptions in this derivation about the initial genotypic frequencies, for example, G_{AA}. The initial adult population does not have to have Hardy–Weinberg genotype frequencies for the zygotes to have Hardy–Weinberg frequencies; all that is required is random mating of the adults regardless of their genotype frequencies. Hence, it takes *only one generation of random mating* to achieve Hardy–Weinberg genotype frequencies regardless of the starting genotype frequencies.

Weinberg's derivation differed from Hardy's at the point of modeling uniting gametes. To Weinberg, random mating meant that the probability of two genotypes being involved in a mating event was simply the product of their respective genotype frequencies. Given a mating, offspring genotypes would be produced according to standard Mendelian probabilities. Hence, in Weinberg's derivation, the mechanisms of producing gametes and the mechanisms of gametic union are utilized in an integrated fashion, as shown in Table 2.3. Note that this table makes an additional assumption not needed under the monoecious version of Hardy, namely, that the genotype frequencies are identical in both sexes. With

Table 2.3. Weinberg's Derivation of Hardy–Weinberg Genotype Frequencies

		Mendelian Probabilities of Offspring (Zygotes)		
Mating Pair	Frequency of Mating Pair	AA	Aa	aa
$AA \times AA$	$G_{AA} \times G_{AA} = G^2_{AA}$	1	0	0
$AA \times Aa$	$G_{AA} \times G_{Aa} = G_{AA}G_{Aa}$	$\frac{1}{2}$	$\frac{1}{2}$	0
$Aa \times AA$	$G_{Aa} \times G_{AA} = G_{AA}G_{Aa}$	$\frac{1}{2}$	$\frac{1}{2}$	0
$AA \times aa$	$G_{AA} \times G_{aa} = G_{AA}G_{aa}$	0	1	0
$aa \times AA$	$G_{aa} \times G_{AA} = G_{AA}G_{aa}$	0	1	0
$Aa \times Aa$	$G_{Aa} \times G_{Aa} = G^2_{Aa}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$Aa \times aa$	$G_{Aa} \times G_{aa} = G_{Aa}G_{aa}$	0	$\frac{1}{2}$	$\frac{1}{2}$
$aa \times Aa$	$G_{aa} \times G_{Aa} = G_{Aa}G_{aa}$	0	$\frac{1}{2}$	$\frac{1}{2}$
$aa \times aa$	$G_{aa} \times G_{aa} = G^2_{aa}$	0	0	1
Total offspring		G'_{AA}	G'_{Aa}	G'_{aa}

Summing zygotes over all mating types:

$G'_{AA} = G^2_{AA} + \frac{1}{2}[2G_{AA}G_{Aa}] + \frac{1}{4}G^2_{Aa} = [G_{AA} + \frac{1}{2}G_{Aa}]^2 = p^2$
$G'_{Aa} = \frac{1}{2}[2G_{AA}G_{Aa}] + 2G_{AA}G_{aa} + \frac{1}{2}G^2_{Aa} + \frac{1}{2}[G_{Aa}G_{aa}] = 2[G_{AA} + \frac{1}{2}G_{Aa}][G_{aa} + \frac{1}{2}G_{Aa}] = 2pq$
$G'_{aa} = \frac{1}{4}G^2_{Aa} + \frac{1}{2}[2G_{Aa}G_{aa}] + G^2_{aa} = [G_{aa} + \frac{1}{2}G_{Aa}]^2 = q^2$

this additional assumption, the end result of Weinberg's derivation is the same as Hardy's: The zygotic genotype frequencies (and hence the adult genotype frequencies of the next generation under the assumptions made here) are again given by $G'_{AA} = p^2$, $G'_{Aa} = 2pq$, and $G'_{aa} = q^2$.

We now address the important question of whether or not microevolution has occurred in this model; that is, are the allele frequencies in the offspring generation different or the same as the allele frequencies of the parent generation. Given that the adults of the offspring generation have the genotype frequencies $G'_{AA} = p^2$, $G'_{Aa} = 2pq$, and $G'_{aa} = q^2$, the allele frequencies in the pool of gametes they produce (say p' for A and q' for a) are calculated from equation 2.1 as

$$p' = p^2 + \frac{1}{2}(2pq) = p^2 + pq = p(p+q) = p \tag{2.2}$$

and $q' = q$ (also shown in Figure 2.1). The allele frequencies p and p' make a contrast at comparable stages in two successive generations (here at the stage of producing gametes), and this contrast allows us to see if evolution has occurred. Because $p = p'$, by definition there has been no evolution. Hence, the Hardy–Weinberg model predicts that allele frequencies are stable over time and that no evolution is occurring under this set of assumptions. Because of this stability over time, Hardy–Weinberg genotype frequencies are often called the Hardy–Weinberg *equilibrium*. As noted earlier, it takes only one generation of random mating to achieve Hardy–Weinberg frequencies, and once achieved the population will remain in this state until one or more assumptions of the Hardy–Weinberg model are violated.

EXAMPLE OF HARDY–WEINBERG LAW

As an illustration of the application of this model, consider a human population of Pueblo Indians scored for genetic variation at the autosomal blood group locus *MN* (Figure 2.3). This locus has two common alleles in most human populations, the *M* allele and the *N* allele. Genetic variation at this locus determines your MN blood group type, with a very simple genotype-to-phenotype mapping: *MM* genotypes have blood group M, *MN* genotypes have blood group MN, and *NN* genotypes have blood group N. Hence, it is easy to characterize the genotypes of all individuals in a population by determining their MN blood group type. Figure 2.3 shows the number of individuals with each of the possible genotypes at this locus in a sample of 140 Pueblo Indians (Boyd 1950). The first step in analyzing a population is to convert the genotype numbers into genotype frequencies by dividing the number of individuals of a given genotype by the total sample size. For example, 83 Pueblo Indians had the *MM* genotype out of the total sample of 140, so the frequency of the *MM* genotype in that sample is $83/140 = 0.593$. Figure 2.3 then shows how the allele frequencies are calculated in the pool of potential gametes, yielding p (the frequency of M in this case) $= 0.757$ and $q = 0.243$. We can also apply the other definition of gene pool to this sample: The gene pool is the population of genes collectively shared by all the individuals. Since this is a diploid locus, the 140 Pueblo Indians collectively share 280 copies of genes at the *MN* locus. The 166 copies found in the 83 *MM* homozygotes are all M, and half of the 92 copies found in the 46 *MN* heterozygotes are M. Hence, the total number of M alleles in this sample of 280 genes is $166 + \frac{1}{2} \times 92 = 212$. The frequency of the M allele is therefore $212/280 = 0.757$. As this shows, either way of conceptualizing the gene pool leads to the same answer.

Continuing with Figure 2.3, we can see that the zygotic frequencies should be 0.573 for *MM*, 0.368 for *MN*, and 0.059 for *NN* if this population were randomly mating. Recall that

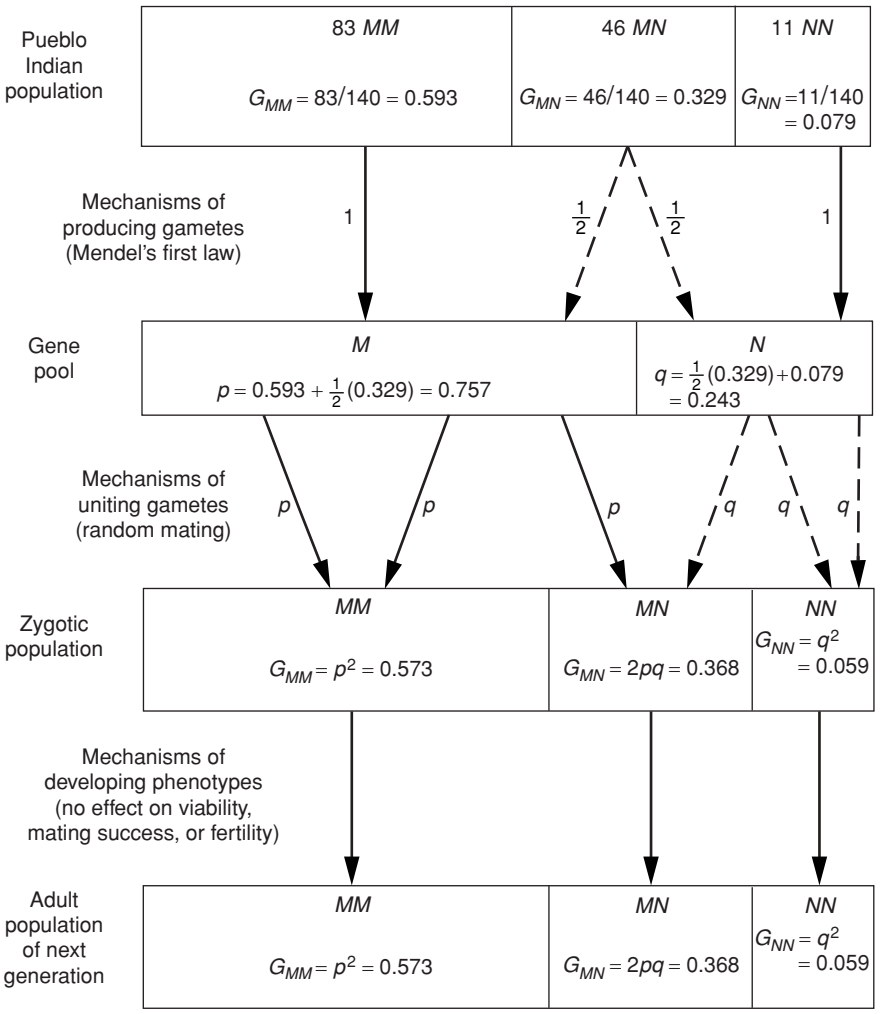

Figure 2.3. Application of Hardy–Weinberg model to sample of Pueblo Indians scored for their genotypes at autosomal *MN* blood group locus.

random mating in this case simply means that the individuals are choosing mates at random with respect to their MN blood group types; it does *not* mean that mating is random for every trait! For example, this population is evenly split between males and females, so the frequency of the female genotype *XX* (where *X* designates the human X chromosome) is 0.5 and the frequency of the male genotype *XY* is 0.5 (where *Y* designates the human Y chromosome). Because sex is determined by the X and Y chromosomes as wholes and these chromosomes do not normally recombine, we effectively can treat gender as determined by a single locus with two alleles, *X* and *Y*. The frequency of *X* gametes in the Pueblo Indian gene pool is $0.5 + \frac{1}{2}(0.5) = 0.75$ and the frequency of *Y* gametes is $\frac{1}{2}(0.5) = 0.25$. Therefore, we would expect the Hardy–Weinberg genotype frequencies of

$$G_{XX} = (0.75)^2 = 0.5625 \qquad G_{XY} = 2(0.75)(0.25) = 0.375$$
$$G_{YY} = (0.25)^2 = 0.0625$$

Obviously, this population is not at Hardy–Weinberg equilibrium for the X and Y chromosomes, and the reason is straightforward: Mating is not random for these genetic elements. Instead, the only cross that can yield offspring is $XX \times XY$, a gross deviation from the Hardy–Weinberg model portrayed in Table 2.3. Because of this highly nonrandom system of mating, the X and Y chromosomes can never achieve Hardy–Weinberg frequencies. Hence, systems of mating can be *locus specific* and Hardy–Weinberg frequencies are only for loci that have a *random system of mating*. Other genetic systems found in the same individuals in the same population may deviate from Hardy–Weinberg because mating is not random for that genetic system.

Recall that when the assumptions of Hardy–Weinberg are met, the population goes to Hardy–Weinberg genotype frequencies in a single generation and remains at those frequencies. Hence, if the Pueblo Indian population had been randomly mating for the MN blood groups in the past and if the other assumptions of Hardy–Weinberg are at least approximately true, we would expect the adult genotype frequencies of the next generation shown in Figure 2.3 to hold for the current adult population as well. This observation provides a basis for testing the hypothesis that this, or any population, has Hardy–Weinberg frequencies. The statistical details and a worked example of such a test are provided in Box 2.1.

IMPORTANCE OF HARDY–WEINBERG LAW

At first, the Hardy–Weinberg law may seem a relatively minor, even trivial, accomplishment. Nevertheless, this simple model played an important role in the development of both genetics and evolution in the early part of the twentieth century. Mendelian genetics had been rediscovered at the start of the twentieth century, but many did not accept it. One of the early proponents of Mendelian genetics was R. C. Punnett (of "Punnett square" fame). Punnett made a presentation at a scientific meeting in which he argued that the trait of brachydactyly (short fingers) was inherited as a Mendelian dominant trait in humans. Udny Yule, a member of the audience, raised the objection that one would expect a 3 : 1 ratio of people with brachydactyly to those without if the Mendelian model were true, and this clearly was not the case. Punnett suspected that there was an error in this argument, but he could not come up with a response at the meeting. Later Punnett explained the problem to his mathematician friend, G. H. Hardy, who immediately proceeded to derive his version of the Hardy–Weinberg law. Hardy's derivation made it clear that Yule had confused the family Mendelian ratio of 3 : 1 (which was for the offspring of a specific mating between two heterozygotes for the dominant trait) with the frequency in a population. Suppose in our earlier derivation that the A allele is dominant over a for some trait. Then the Hardy–Weinberg law predicts that the ratio of frequencies of those with the dominant trait to those with the recessive trait in a random-mating population should be $p^2 + 2pq : q^2$. There is no constraint upon this ratio to be 3 : 1 or any of the other family ratios expected under Mendelian inheritance. Rather, this population ratio can vary continuously as p varies from 0 to 1.

The predicted ratio of individuals with dominant to recessive traits also provided a method for predicting the frequency of carriers for genetic disease. Many genetic diseases in humans are recessive, so now let a be a recessive disease allele. Only two phenotypic categories could be observed in these early Mendelian studies: the dominant phenotype, associated with the genotypes AA and Aa, and the recessive, associated with the genotype aa. Thus, there was seemingly no way to predict how many people were carriers (Aa) as

BOX 2.1 TESTING TO SEE IF A POPULATION IS IN SINGLE-LOCUS HARDY–WEINBERG

We first estimate the allele frequencies using either equation 2.1 or the gene-counting method and then calculate the expected Hardy–Weinberg genotype frequencies. These steps have already been done for the Pueblo Indians, as shown in Figure 2.3. Next, we convert the expected Hardy–Weinberg genotype frequencies into expected genotype numbers by multiplying each frequency by the total sample size, which is 140 in this case. For example, the expected number of *MM* homozygotes under Hardy–Weinberg for the Pueblo Indian sample is $0.573 \times 140 = 80.22$. Similarly, the expected numbers of *MN* and *NN* genotypes are 51.52 and 8.26, respectively. Now we can calculate a standard chi-square statistic (see Appendix 2):

$$\sum_{\text{genotypes}} \frac{\left[\text{Obs}(i) - \text{Exp}(i)\right]^2}{\text{Exp}(i)} = \frac{(83 - 80.22)^2}{80.22} + \frac{(46 - 51.52)^2}{51.52} + \frac{(11 - 8.26)^2}{8.26} = 1.59 \tag{2.3}$$

where Obs(i) is the observed number of individuals with genotype i and Exp(i) is the expected number of individuals with genotype i under Hardy–Weinberg (in this case i can be *MM*, *MN*, or *NN*). If the null hypothesis of Hardy–Weinberg is true, we expect the statistic calculated in equation 2.3 to have a value such that there is a high probability of the statistic having that or a higher value when in fact the population is at Hardy–Weinberg. To calculate this probability, we need the degrees of freedom associated with the chi-square statistic. In general, the degree of freedom is the number of categories being tested (three genotype categories in this case) minus 1 minus the number of independent parameters that had to be estimated from the data being tested to generate the expected numbers. In order to generate the Hardy–Weinberg expected values, we first had to estimate the allele frequencies of *M* and *N* from the data being tested. However, recall that $q = 1 - p$, so that once we know p, we automatically know q. This means that the data are used to estimate only one independent parameter (the parameter p). Therefore, the degree of freedom is $3 - 1 - 1 = 1$. We can now look up the value of 1.59 with one degree of freedom in a chi-square table or statistical calculator and find that the probability of getting a value of 1.59 or larger if the null hypothesis of Hardy–Weinberg were true is 0.21. Generally, such probabilities have to be less than 0.05 before the null hypothesis is rejected. Hence, we fail to reject the null hypothesis of Hardy–Weinberg for this sample of Pueblo Indians scored for the *MN* locus. It would have been simpler to say that the Pueblo Indian population is in Hardy–Weinberg, but we have not actually demonstrated this. Our sample is relatively small, and perhaps with more extensive sampling we would reject Hardy–Weinberg. Hence, all that we have really demonstrated is that we fail to reject Hardy–Weinberg for our current sample. Statistical tests never prove that a null hypothesis is true; the test either rejects or fails to reject the null hypothesis.

they could not be distinguished phenotypically from the *AA* homozygotes. However, if we assume Hardy–Weinberg is true, then the frequency of individuals affected with the genetic disease (which is observable) is q^2. Hence, we can estimate q is this case as

$$\hat{q} = \sqrt{G_{aa}} \tag{2.4}$$

Given $\hat{q}$, the frequency of carriers of the genetic disease can be estimated as $2(1-\hat{q})\hat{q}$. Note, in this case, we cannot actually test the population for Hardy–Weinberg because we only have two observable categories and we have estimated one parameter from the data to be tested (equation 2.3). Therefore, the degrees of freedom are $2-1-1=0$. Zero degrees of freedom means we have insufficient information in the data to test the model (Appendix 2). Equation 2.4 should never be used when all genotypic classes are observable because it is valid only in the special case of Hardy–Weinberg genotype frequencies. In contrast, equation 2.1 makes no assumptions about Hardy–Weinberg and is true for any set of genotype frequencies. Therefore, when all genotypic classes are observable, equation 2.1 should be used instead of equation 2.4 because equation 2.1 will always give you the right answer whereas equation 2.4 will only give the right answer in a specific special case. Nevertheless, equation 2.4 played an important role throughout much of the twentieth century in genetic counseling in predicting heterozygous carrier frequencies for autosomal recessive genetic diseases when all genotypic classes were not observable.

The Hardy–Weinberg law also predicts no evolution; that is, the allele frequencies remain constant over time. At first this may also seem to be a rather uninteresting result, but this observation was critical for the acceptance of the Darwin–Wallace concept of natural selection. The publication of Darwin's book *The Origin of Species* in 1859 strongly established the concept of descent with modification within biology. However, Darwin's (and Wallace's) explanation for the origin of adaptations via natural selection was less universally accepted. Darwin felt that the Scottish engineer Fleeming Jenkin raised one of the most serious objections to the theory of natural selection in 1867. At this time, the dominant idea of inheritance was that of "blending inheritance" in which the traits of the father and mother are blended together, much as mixing two different colors of paint together results in a new color that represents equal amounts of the original colors. Jenkin pointed out that half of the heritable variation would be lost every generation under blending inheritance; hence, a population should quickly become homogeneous. Recall from Chapter 1 that heritable variation is a necessary prerequisite for all evolution, so evolution itself would grind to a halt unless mutation replenished this loss at the same rate. Darwin and Wallace had based their theories of natural selection upon the tenet that mutation creates new variation at random with respect to the needs of the organism in coping with its environment. It seemed implausible that half of the genetic material could mutate at random every generation and the organisms still survive. Hence, Jenkin's argument seemed to imply that either genetic variation would quickly vanish and all evolution halt or that natural selection required levels of mutation that would result in extinction. This problem even led Darwin in his 1868 book *The Variation of Animals and Plants under Domestication* to speculate that mutation might be directed by the environment. By the beginning of the twentieth century, many neo-Lamarkian ideas based upon directed mutations were popular alternatives to natural selection of random mutations.

Jenkin's argument was finally put to rest by the Hardy–Weinberg law. The Hardy–Weinberg model, by ignoring many potential evolutionary forces (Table 2.1), focuses our attention upon the potential evolutionary impact of Mendelian inheritance alone. By demonstrating that Mendelian inheritance results in a population with a constant allele frequency, it was evident that Mendelian genetic variation is not rapidly lost from a population. Indeed, under the strict assumptions of Hardy–Weinberg, genetic variation persists indefinitely. Thus, even though the Hardy–Weinberg model is one of no evolution, this model was critical for the acceptance of natural selection as a plausible mechanism of evolutionary change under Mendelian inheritance.

In general, this book is concerned about evolutionary change. In modeling evolution, Hardy–Weinberg is a useful null model of evolutionary stasis. Indeed, much of the rest of this book is devoted to relaxing one or more of the assumptions of the original Hardy–Weinberg model and seeing whether or not evolution can result. In this sense, Hardy–Weinberg serves as a valuable springboard for the investigation of many forces of evolutionary change. In the remainder of this chapter we consider just one slight deviation from the original Hardy–Weinberg model, and we will investigate the evolutionary implications of this slight change.

HARDY–WEINBERG FOR TWO LOCI

The original Hardy–Weinberg model assumed a genetic architecture of one autosomal locus with two alleles. We will now consider a slightly more complicated genetic architecture of two autosomal loci, each with two alleles (say A and a at locus 1 and B and b at locus 2). Otherwise, we will retain all other assumptions of the original Hardy–Weinberg model. However, there is one new assumption. Recall from Chapter 1 that our second premise is that DNA can mutate and recombine. We will retain the Hardy–Weinberg assumption of no mutation, but we will allow recombination (either independent assortment if the two loci are on different autosomes or crossing over if they are on the same autosome).

Because our main interest is on whether or not evolutionary change occurs, we will start with the gene pool and go to the next generation's gene pool (Figure 2.4), rather than going from adult population to adult population as in Figures 2.1 and 2.3. Given two loci with two alleles each and the possibility of recombination between them, a total of four gamete types are possible (*AB, Ab, aB,* and *ab*). The gene pool is characterized by four gamete frequencies (Figure 2.4), symbolized by g_{xy}, where x indicates the allele at locus 1 and y indicates the allele at locus 2. Just as p and q sum to 1, these four gamete frequencies also sum to 1 because they define a probability distribution over the gene pool. The transition from this gene pool to the zygotes is governed by the same population structure (rules of uniting gametes) as given in the single-locus Hardy–Weinberg. In particular, the assumption of random mating means that gametes are drawn independently from the gene pool, with the probability of any given gamete type being equal to its frequency. The probability of any particular genotype is simply the product of its gamete frequencies, just as in the single-locus Hardy–Weinberg model. In Figure 2.4 we are not keeping track of the paternal or maternal origins of any gamete, so both types of heterozygotes are always pooled and therefore the product of the gamete frequencies for heterozygous genotypes is multiplied by 2. For example, the frequency of the genotype *AB*/*Ab* is $2g_{AB}g_{Ab}$. Note that there are two types of double heterozygotes, *AB*/*ab* (the cis double heterozygote with a random-mating frequency of $2g_{AB}g_{ab}$) and *Ab*/*aB* (the trans double-heterozygote with a random-mating frequency of $2g_{Ab}g_{aB}$). Although the cis and trans double heterozygotes share the double-heterozygous genotype, completely different gamete types produce the cis and trans double-heterozygosity. As we will soon see, the cis and trans double-heterozygous genotypes contribute to the gene pool in different ways. Hence, we will keep the cis and trans double-heterozygote classes separate.

The rules for uniting gametes in the two-locus model are the same as for the single-locus model, the only difference being that there are now 10 genotypic combinations. As with the single-locus model, if we know the gamete frequencies and know that the mating is at random (along with the other population structure assumptions of Hardy–Weinberg),

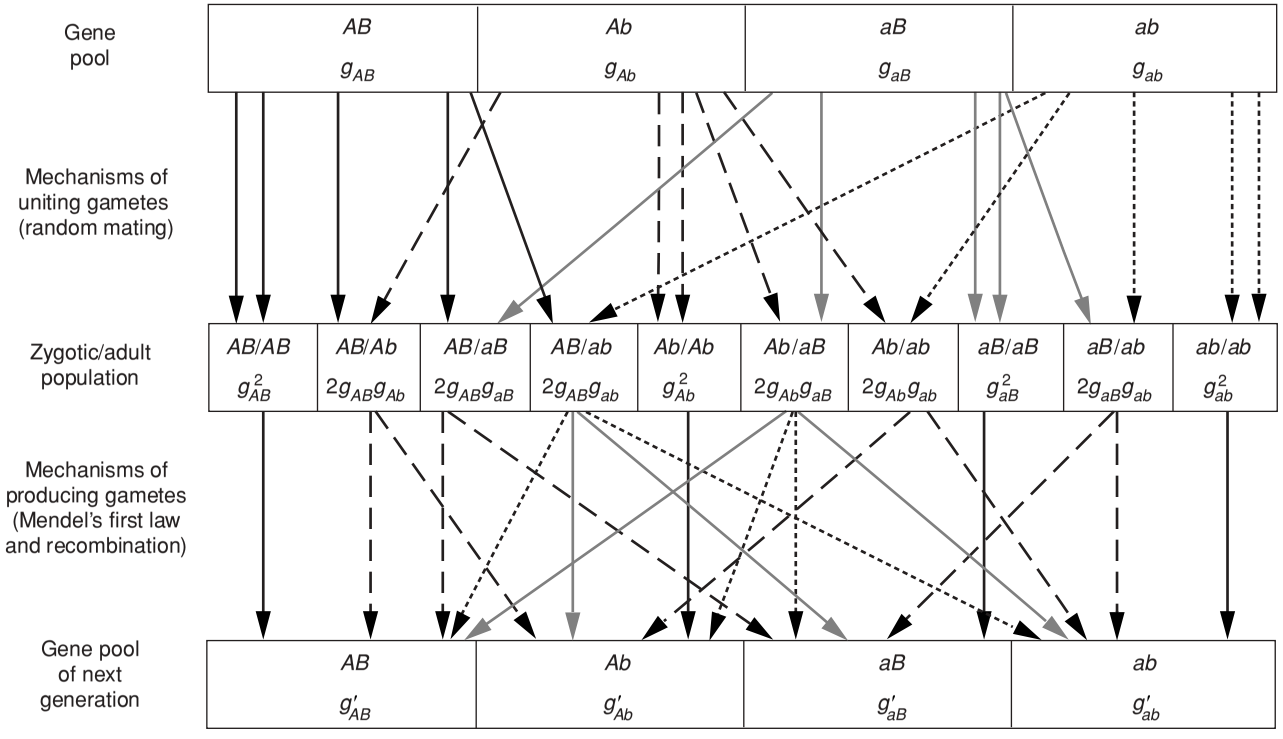

Figure 2.4. Derivation of Hardy–Weinberg law for two autosomal loci with two alleles each: *A* and *a* at locus 1 and *B* and *b* at locus 2.In going from gametes to zygotes, solid lines represent gametes bearing the *AB* alleles and are assigned the weight g_{AB}, dashed lines represent gametes bearing the *Ab* allele and are assigned the weight g_{Ab}, grey lines represent gametes bearing the *aB* gametes and are assigned the weight g_{aB}, and dotted lines represent gametes bearing the *ab* alleles and are assigned the weight g_{ab}. In going from adults to gametes, solid lines represent Mendelian transition probabilities of 1 for homozygotes, dashed lines represent Mendelian transition probabilities of $\frac{1}{2}$ for single heterozygotes, dotted lines represent nonrecombinant Mendelian transition probabilities of $\frac{1}{2}(1-r)$ (where r is the recombination frequency between loci 1 and 2) for double heterozygotes, and grey lines represent recombinant Mendelian transition probabilities of $\frac{1}{2}r$ for double heterozygotes.

we can predict the zygotic genotype frequencies. If we further assume that there are no phenotypic differences that affect viability, mating success, or fertility, we can also predict the next generation's adult genotype frequencies from the gamete frequencies.

The similarities to the single-locus model end when we advance to the transition from the next generation's adult population to the gene pool of the next generation (Figure 2.4). At this point, some new rules are encountered in producing gametes that did not exist at all in the single-locus model (Figure 2.1). As before, homozygous genotypes can only produce gametes bearing the alleles for which they are homozygous (this comes from the assumptions of normal meiosis and no mutations). As before, genotypes heterozygous for just one locus produce two gamete types, with equal frequency as stipulated by Mendel's first law. However, genotypes that are heterozygous for both loci can produce all four gamete types, and the probabilities are determined by a combination of Mendel's first law and recombination (Mendel's second law of independent assortment if the loci are on different chromosomes or the recombination frequency if on the same chromosome). Hence, the transition from genotypes to gametes requires a new parameter, the recombination frequency r, which is $\frac{1}{2}$ if the loci are on different chromosomes and $0 \leq r \leq \frac{1}{2}$ if the loci are on the same chromosome.

The addition of recombination produces some qualitative differences with the single-locus model. First, in the single-locus model, an individual could only pass on gametes of the same types that the individual inherited from its parents. But note from Figure 2.4 that the cis double heterozygote *AB*/*ab*, which inherited the cis *AB* and *ab* gamete types from its parents, can produce not only the cis gamete types, each with probability $\frac{1}{2}(1-r)$, but also the trans gamete types *Ab* and *aB*, each with probability $\frac{1}{2}r$. Similarly, the trans double heterozygote can produce both cis and trans gamete types (Figure 2.4). Thus, recombination allows the double heterozygotes to produce gamete types that they themselves did not inherit from their parents. This effect of recombination is found only in the double-heterozygote class, but this does not mean that recombination only occurs in double heterozygotes. Consider, for example, the single heterozygote *AB*/*Ab*. If no recombination occurs in meiosis, this genotype will produce the gamete types *AB* and *Ab* with equal frequency. Hence, the total probability of gamete type *AB* with no recombination is $\frac{1}{2}(1-r)$, and similarly it is $\frac{1}{2}(1-r)$ for *Ab*. Now consider a meiotic event in which recombination did occur. Such a recombinant meiosis also produces the gamete types *AB* and *Ab* with equal frequency, that is, with probability $\frac{1}{2}r$ for each. However, in the recombinant *AB* gamete the *A* allele that is combined with the *B* allele originally came from the *Ab* gamete that the *AB*/*Ab* individual inherited from one of its parents. Hence, recombination has occurred, but because we do not distinguish among copies of the *A* alleles, we see no observable genetic impact. Hence, the total probability of an *AB* gamete, regardless of the source of the *A* allele, is $\frac{1}{2}(1-r)+\frac{1}{2}r=\frac{1}{2}$, and the total probability of an *Ab* gamete, regardless of the source of the *A* allele, is $\frac{1}{2}(1-r)+\frac{1}{2}r=\frac{1}{2}$. Thus, recombination is occurring in all genotypes but is observable only in double heterozygotes.

The qualitative difference from the single-locus model that causes some genotypes to produce gamete types that they themselves did not inherit leads to yet another qualitative difference: The two definitions of gene pool given in Chapter 1 are no longer equivalent. If we define the gene pool as the shared genes of all the adult individuals, we obtain the gamete frequencies from the pool of gametes produced by their parents (the g_{xy}'s in Figure 2.4). On the other hand, if we define the gene pool as the population of potential gametes produced by all the adult individuals, the effects of recombination enter and we obtain the g'_{xy}'s in Figure 2.4. To avoid any further confusion on this point, the term "gene pool" in this book

will always refer to the population of potential gametes unless otherwise stated. The general population genetic literature often does not make this distinction because in the standard single-locus Hardy–Weinberg model it is not important. Quite frequently there is a time difference of one generation among the models of various authors depending upon which definition of gene pool they use (usually implicitly). Therefore, readers have to be careful in interpreting what various authors mean by gene pool when dealing with multilocus models or other models in which these two definitions may diverge.

The most important qualitative difference from the single-locus model involves the potential for evolution. As seen before, the single-locus Hardy–Weinberg model goes to equilibrium in a single generation of random mating and then stays at equilibrium, resulting in no evolution. To see if this is the case for the two-locus model, we now use equation 2.1 to calculate the gamete frequency of the *AB* gamete using the weights implied by the arrows in Figure 2.4 going from adults to gametes:

$$\begin{aligned} g'_{AB} &= 1 \cdot g_{AB}^2 + \frac{1}{2}(2g_{AB}g_{Ab}) + \frac{1}{2}(2g_{AB}g_{aB}) + \frac{1}{2}(1-r)(2g_{AB}g_{ab}) + \frac{1}{2}r(2g_{Ab}g_{aB}) \\ &= g_{AB}\left[g_{AB} + g_{Ab} + g_{aB} + (1-r)g_{ab}\right] + r g_{Ab}g_{aB} \\ &= g_{AB}\left[g_{AB} + g_{Ab} + g_{aB} + g_{ab}\right] + r g_{Ab}g_{aB} - r g_{AB}g_{ab} \\ &= g_{AB} + r(g_{Ab}g_{aB} - g_{AB}g_{ab}) = g_{AB} - rD \end{aligned} \tag{2.5}$$

where $D = (g_{AB}g_{ab} - g_{Ab}g_{aB})$. The parameter D is commonly known as **linkage disequilibrium**. However, because it can exist for pairs of loci on different chromosomes that are not linked at all, a more accurate but more cumbersome term is **gametic-phase imbalance**. Because the term linkage disequilibrium dominates the literature, we will use it throughout the book, but with the caveat that it can be applied to unlinked loci.

Similarly, the other three gamete types can be obtained from equation 2.1 as

$$\begin{aligned} g'_{Ab} &= 1 \cdot g_{Ab}^2 + \frac{1}{2}(2g_{AB}g_{Ab}) + \frac{1}{2}(2g_{Ab}g_{ab}) + \frac{1}{2}(1-r)(2g_{Ab}g_{aB}) + \frac{1}{2}r(2g_{AB}g_{ab}) \\ &= g_{Ab} + rD \\ g'_{aB} &= 1 \cdot g_{aB}^2 + \frac{1}{2}(2g_{AB}g_{aB}) + \frac{1}{2}(2g_{aB}g_{ab}) + \frac{1}{2}(1-r)(2g_{Ab}g_{aB}) + \frac{1}{2}r(2g_{AB}g_{ab}) \\ &= g_{aB} + rD \\ g'_{ab} &= 1 \cdot g_{ab}^2 + \frac{1}{2}(2g_{Ab}g_{ab}) + \frac{1}{2}(2g_{aB}g_{ab}) + \frac{1}{2}(1-r)(2g_{AB}g_{ab}) + \frac{1}{2}r(2g_{Ab}g_{aB}) \\ &= g_{ab} - rD \end{aligned} \tag{2.6}$$

At this point, we can now address our primary question: Is evolution occurring? Recall our definition from Chapter 1 of evolution as a change in the frequencies of various types of genes or gene combinations in the gene pool. As is evident from equations 2.5 and 2.6, as long as $r > 0$ (that is, some recombination is occurring) and $D \neq 0$ (there is some linkage disequilibrium), $g_{xy} \neq g'_{xy}$: *Evolution is occurring*! Thus, a seemingly minor change from one to two loci results in a major qualitative change of population-level attributes.

No evolution occurs in this model if $r = 0$. In that case, the two-locus model is equivalent to a single-locus model with four possible alleles. Thus, some multilocus systems can be treated as if they were a single locus as long as there is no recombination. On the other hand, recombination can sometimes occur *within* a single gene. As mentioned in

Chapter 1, the genetic variation within a 9.7-kb segment of the *lipoprotein lipase* (*LPL*) gene in humans was shaped in part by about 30 recombination events (Templeton et al. 2000a). Thus, in some cases the evolutionary potential created by recombination must be considered even at the single-locus level. In the case of *LPL*, we are looking at two or more different polymorphic nucleotide sites within the same gene and not, technically speaking, at different loci. However, the qualitative evolutionary potential is still the same as long as the polymorphic sites under examination can recombine, regardless of whether those sites are single nucleotides within a gene or traditional loci.

No evolution also occurs in this model if $D = 0$. Here, D will equal zero when the two-locus gamete frequencies are the product of their respective single-locus allele frequencies. To see this, let p_A be the frequency of the A allele at locus 1 and p_B the frequency of the B allele at locus 2. These single-locus allele frequencies are related to the two-locus gamete frequencies by

$$p_A = g_{AB} + g_{Ab} \qquad p_B = g_{AB} + g_{aB} \tag{2.7}$$

Now consider the product of the A and B allele frequencies:

$$\begin{aligned} p_A p_B &= (g_{AB} + g_{Ab})(g_{AB} + g_{aB}) \\ &= g_{AB}^2 + g_{AB}g_{aB} + g_{AB}g_{Ab} + g_{Ab}g_{aB} \\ &= g_{AB}(g_{AB} + g_{aB} + g_{Ab}) + g_{Ab}g_{aB} \\ &= g_{AB}(1 - g_{ab}) + g_{Ab}g_{aB} \\ &= g_{AB} - g_{AB}g_{ab} + g_{Ab}g_{aB} \\ &= g_{AB} - D \end{aligned} \tag{2.8}$$

Solving equation 2.8 for D yields

$$D = g_{AB} - p_A p_B \tag{2.9}$$

and similar equations can be derived in terms of the other three gamete frequencies. Equation 2.9 suggests another biological interpretation of D; *it is the deviation of the two-locus gamete frequencies from the product of the respective single-locus allele frequencies*. Equation 2.9 also makes it clear that D will be zero when the two-locus gamete frequency is given by the product of the respective single-locus allele frequencies. This can also be seen by evaluating the original formula for linkage disequilibrium under the assumption that the two-locus gamete frequencies are given the product of their respective allele frequencies: $D = g_{AB}g_{ab} - g_{Ab}g_{aB} = (p_A p_B)(p_a p_b) - (p_A p_b)(p_a p_B) = p_A p_B p_a p_b - p_A p_b p_a p_B = 0$.

The two-locus gamete frequencies will be products of the single-locus allele frequencies when knowing what allele is present at one locus in a gamete does not alter the probabilities of the alleles at the second locus; that is, the probabilities of the alleles at the second locus are simply their respective allele frequencies regardless of what allele occurs at the first locus. When $D \neq 0$, knowing which allele a gamete bears at one locus does influence the probabilities of the alleles at the second locus. In statistical terms, $D = 0$ means that there is no association in the population between variation at locus 1 with variation at locus 2. When $D = 0$, equations 2.5 and 2.6 show that the gamete frequencies (and hence the genotype

frequencies) are constant, just as they were in the single-locus Hardy–Weinberg model. Thus, when $D = 0$ the population is at a nonevolving equilibrium, given the other standard Hardy–Weinberg assumptions. We can now understand why D is called *disequilibrium*. When D is not zero and there is recombination, the population is evolving and is not at a two-locus Hardy–Weinberg equilibrium. The larger D is in magnitude, the greater this deviation from two-locus equilibrium.

Evolution occurs when $r > 0$ and $D \neq 0$, and we now examine the evolutionary process induced by linkage disequilibrium in more detail. From Figure 2.4 or equations 2.5 and 2.6, we see that linkage disequilibrium in the original gene pool ($g_{AB}g_{ab} - g_{Ab}g_{aB}$) influences the next generation's gene pool. Similarly, the linkage disequilibrium in the next generation's gene pool will influence the subsequent generation's gene pool. The linkage disequilibrium in the next generation's gene pool in Figure 2.4 is

$$\begin{aligned} D_1 &= [g'_{AB}g'_{ab} - g'_{aB}g'_{Ab}] \\ &= [(g_{AB} - rD)(g_{ab} - rD) - (g_{aB} + rD)(g_{Ab} + rD)] \\ &= D(1 - r) \end{aligned} \tag{2.10}$$

Using equation 2.10 recursively, we can see that D_2 (the linkage disequilibrium in the gene pool two generations removed from the original gene pool) is $D(1 - r)^2$. In general, if we start with some initial linkage disequilibrium, say D_0, then D_t, the linkage disequilibrium after t generations of random mating, is

$$D_t = D_0(1 - r)^t. \tag{2.11}$$

Equation 2.11 reveals that the evolution induced by linkage disequilibrium is both gradual and directional, as illustrated in Figure 2.5. Because $r \leq \frac{1}{2}$, the quantity $(1 - r)^t$ goes to

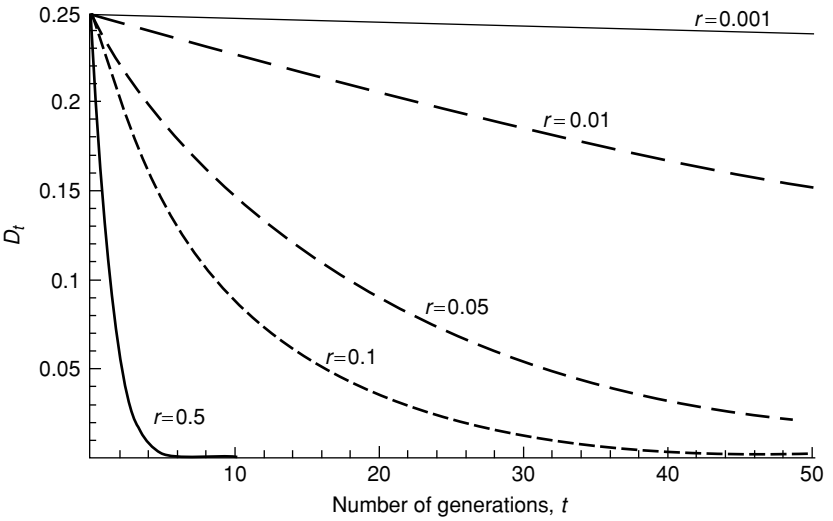

Figure 2.5. Decay of linkage disequilibrium with time in generations as function of different recombination rates r starting with initial value of $D_0 = 0.25$.

zero as the number of generations (t) gets large. Hence, the direction of evolution is to dissipate linkage disequilibrium and to move closer and closer to a two-locus Hardy–Weinberg equilibrium in which the two-locus gamete frequencies are the products of the constituent single-locus allele frequencies. The approach to this equilibrium is gradual, proceeding at an exponential rate determined by $1 - r$. The larger the value of r (i.e., the more recombination occurs), the faster is the approach to equilibrium with $D = 0$ (Figure 2.5). Note, however, that even for loci on different chromosomes that sort independently ($r = \frac{1}{2}$), equilibrium is not attained instantly (Figure 2.5), in great contrast to the single-locus Hardy–Weinberg model. However, the approach to linkage equilibrium is quite rapid with unlinked loci. For example, after just five generations of random mating, only a little more than 3% of the original disequilibrium for unlinked loci remains (from equation 2.8). However, for linked loci with r small, the linkage disequilibrium can persist and affect gene pool evolution for many, many generations. For example, for two loci with $r = 0.01$ (1% recombination), it takes 345 generations to reduce the initial linkage disequilibrium to the level achieved in just 5 generations for unlinked loci. During this approach to linkage equilibrium, the two single-locus systems that contribute to the two-locus genetic architecture will be at a single-locus Hardy–Weinberg in just one generation of random mating, but the multilocus system will be in disequilibrium and evolving (given initial linkage disequilibrium). Therefore, recombination and linkage disequilibrium are sufficient conditions for evolution in a multilocus system.

SOURCES OF LINKAGE DISEQUILIBRIUM

Given that some initial linkage disequilibrium is necessary before recombination can act as an evolutionary force in a random-mating population, it is important to understand what factors can create an initial disequilibrium. Many factors can create linkage disequilibrium, including:

- Mutation
- Nonrandom mating
- Finite population size
- Gene flow
- Natural selection

Note that this list of factors that can generate linkage disequilibrium corresponds to the very same factors that are assumed not to occur in the simple Hardy–Weinberg model (Table 2.1).

All of these factors will be considered in this book, but for now we focus only upon the first: mutation. The impact of mutation is most easily seen by considering another measure of linkage disequilibrium known as the **normalized linkage disequilibrium D'**, which is the linkage disequilibrium divided by its theoretical maximum absolute value. Because two-locus gamete frequencies cannot be negative or greater than the corresponding single-locus allele frequencies (equation 2.7), we have that

$$0 \leq g_{AB} \leq \min(p_A, p_B) \tag{2.12}$$

Solving equation 2.9 for g_{AB} and substituting the result into inequality 2.12 yield

$$\begin{aligned} -p_A p_B \le D &\le \min(p_A - p_A p_B, p_B - p_A p_B) \\ &= \min(p_A(1-p_B), p_B(1-p_A)) \qquad \text{or} \qquad -p_A p_B \le D \le \min(p_A p_b, p_a p_B) \end{aligned} \tag{2.13}$$

As noted above, equations similar to 2.9 can be derived with respect to the other gamete frequencies, such as g_{ab}, so D also satisfies the inequality

$$-p_a p_b \le D \le \min(p_A p_b, p_a p_B) \tag{2.14}$$

Thus we have

$$D' = \begin{cases} \dfrac{D}{\min(p_A p_B, p_a p_b)} & D < 0 \\ \dfrac{D}{\min(p_A p_b, p_a p_B)} & D > 0 \end{cases} \tag{2.15}$$

where D' has a range of values from -1 to $+1$ for all pairs of loci irrespective of the allele frequencies at the component loci (although D' itself is still dependent on allele frequencies). This is a great advantage when dealing with the impact of mutation on linkage disequilibrium. When mutations first occur, they are inevitably rare, and this places severe constraints on the magnitude of D (inequalities 2.13 and 2.14). In contrast, D' or $|D'|$ is a better vehicle for investigating the impact of mutation upon linkage disequilibrium because these measures are not constrained in their range of values by the rarity of a new mutation.

Variation is needed at both loci before there is even any potential for linkage disequilibrium. Let us start with a population that has genetic variation only at the A locus, with two alleles A and a at frequencies p_A and p_a, respectively. Suppose that there is no variation at the B locus, with all copies of the B gene being of allelic state B ($p_B = 1$). Therefore, the initial gene pool can be described by the two-locus gamete frequencies $g_{AB} = p_A$ and $g_{aB} = p_a$. Now suppose that a mutational event occurs in one of the copies of the B allele to create the b allele. This initial mutational event must occur either in an A-bearing gamete or in an a-bearing gamete. Suppose this initial mutation occurred in an a-bearing gamete. Such a mutation produces a third gamete type, ab. Let the initial frequency of this gamete type be g_{ab}, which will normally be close to zero because of the recentness of the B-to-b mutation. Note that the fourth potential type of gamete, Ab, does not exist in this gene pool at all because, by assumption, the B-to-b mutation occurred on an a-bearing gamete. Hence, after mutation the gene pool is characterized by $g_{AB} > 0$, $g_{aB} > 0$, $g_{ab} > 0$ but close to 0, and $g_{Ab} = 0$. The initial linkage disequilibrium after mutation has created the third gamete type is therefore $D = g_{AB} g_{ab}$. At the one-locus level, the gene pool after mutation has $p_A = g_{AB}$, $p_a = g_{aB} + g_{ab}$, $p_B = g_{AB} + g_{aB}$, and $p_b = g_{ab}$. Therefore, after mutation $D = p_A p_b$ and $D' = 1$. The other alternative is when the B-to-b mutation occurred on an A-bearing gamete, in which case mutation creates the Ab gamete, and the gene pool after mutation has $g_{AB} > 0$, $g_{aB} > 0$, $g_{Ab} > 0$ but close to 0, and $g_{ab} = 0$. The initial linkage disequilibrium after mutation is now $D = -g_{aB} g_{Ab} = -p_a p_b$ and $D' = -1$. Regardless of whether the b mutation occurs on an a or A background, there will always be an initial linkage disequilibrium that is maximal in magnitude. Hence, the very act of mutation creates

maximal linkage disequilibrium, so multilocus genetic systems *always* begin with linkage disequilibrium. As shown by equation 2.11, this initial linkage disequilibrium decays only gradually with time and extremely slowly when recombination is rare. As a consequence, the linkage disequilibrium created by the act of mutation can sometimes persist for long periods of time after the original mutational event. This observation leads to an important conclusion about evolution: *The current state of a population's gene pool and its ongoing evolution are influenced by its past history.* The past cannot be ignored in understanding the present and predicting the future for biological systems subject to evolutionary change.

SOME IMPLICATIONS OF IMPACT OF EVOLUTIONARY HISTORY UPON DISEQUILIBRIUM

The impact of the past as measured by D_0 upon the present in multilocus systems can be either a boon or a bane, depending upon the question being addressed. As a boon, multilocus studies inherently contain information about the past history of present-day genetic variation and its mutational origins. Parts of this history can often be inferred from multilocus (or multi-nucleotide-site) studies and hence give us a window into the past. For example, one important mutation in human genetics is the sickle cell mutation in the sixth codon of the autosomal locus that codes for the β chain of adult hemoglobin. We will look at the phenotypic and adaptive significance of this mutation later in this book. For now we focus on the linkage disequilibrium patterns of this relatively new mutation in the human gene pool with some genetic variation in surrounding loci. Figure 2.6 shows the genetic state of some of these surrounding loci on chromosomes that contain the β^S allele. As will be detailed later in this book, the β^S allele only recently became common in specific,

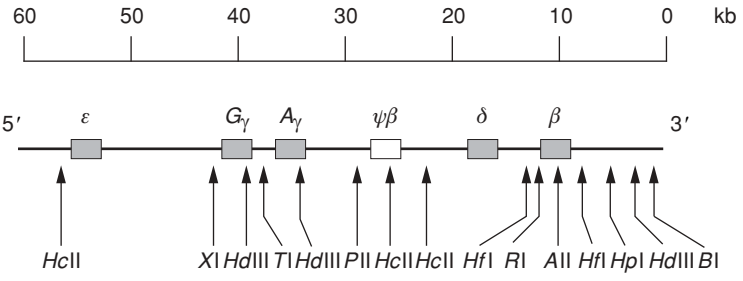

Figure 2.6. Multilocus genetic backgrounds containing β^S alleles at hemoglobin β-chain locus. Restriction site polymorphisms in or near several hemoglobin chain loci ($\beta, \varepsilon, \delta$, $G\gamma$, and $A\gamma$) and the pseudogene $\psi\beta$ are indicated, with "+," meaning that the indicated restriction enzyme cuts the site on that chromosomal type, and "−," meaning that it does not cut. Data from Lapoumeroulie et al. (1992) and Oner et al. (1992).

geographically restricted human populations and globally is a rare allele. In contrast, the restriction site polymorphisms (see Appendix 1) at nearby loci and in noncoding regions between these loci are more widespread and common in human populations. As will be discussed in Chapter 7, this pattern implies that the β^S allele is more recent in origin than the other genetic variants shown in Figure 2.6. As expected for a relatively new mutation in a DNA region showing only low levels of recombination, the β^S allele shows extensive linkage disequilibrium with these restriction site polymorphisms. However, the β^S allele is found on not just one but at least five distinct chromosome backgrounds as defined by multiple-restriction-site polymorphisms (Figure 2.6). This implies that the mutation at the sixth codon that defines the β^S allele occurred multiple times in recent human evolution, at least four times in Africa and one in Asia (Lapoumeroulie et al. 1992; Oner et al. 1992). Hence, by looking at patterns of linkage disequilibrium, we can make inferences about the mutational history of this particular allele. As will be shown in Chapter 7, techniques exist for extracting much historical information when linkage disequilibrium is present, information that can be used to test a number of hypotheses about the history of the locus under study.

As a bane, the lasting evolutionary effects of the initial linkage disequilibrium created by mutation can also limit our abilities to make inference. When recombination rates are on the order of mutation rates or less, then r has little meaning as a "rate" because recombination is a rare and sporadic event that may not have occurred even once in the subsequent evolutionary history of the mutational gene lineage under study. Under these conditions, D'_0 can persist with little or no diminution in magnitude for thousands of generations or perhaps throughout the entire existence of the species. The condition of extremely low rates of recombination is now commonly encountered in population genetic studies because increasingly polymorphisms are scored by sequencing or restriction site mapping of small stretches of DNA. These studies often reveal many polymorphic nucleotide sites in these small DNA regions, as mentioned before with the *LPL* example. As with that example, we are no longer talking about linkage disequilibrium between Mendelian loci, but rather specific sites in the DNA, often within the same gene. Equation 2.11 emphasizes that linkage disequilibrium dissipates with increased recombination. This has led to an expectation that linkage disequilibrium is an inverse surrogate for recombination; that is, the higher the magnitude of linkage disequilibrium, the closer two markers must be on the chromosome. But when recombination is rare and sporadic, the magnitude and pattern of linkage disequilibrium within the DNA region primarily reflect D_0 in equation 2.11 and not r. Thus, the pattern and amount of linkage disequilibrium within small regions of DNA usually are more a reflection of history than of recombination. As a consequence, it is not surprising that many studies of linkage disequilibrium have found little to no correlation between the magnitude of D or D' and the physical distance on the DNA molecule between polymorphic sites in small DNA regions (Templeton 1999a).

For example, Figure 2.7 shows the pattern of statistically significant linkage disequilibrium found within a small autosomal region of the human genome that contains three loci coding for different apoproteins (these proteins combine with lipids and other fats to form apolipoproteins, which can then be transported in the blood stream). Note that the two polymorphic sites that are in closest physical proximity are an insertion/deletion polymorphism and a nearby polymorphic *Xmn*I restriction site. The two most distant sites in this study are the insertion/deletion polymorphic site and a polymorphic *Pvu*II restriction site. Yet, there is no significant linkage disequilibrium between the insertion/deletion site and the *Xmn*I site, whereas there is significant disequilibrium between the insertion/deletion site

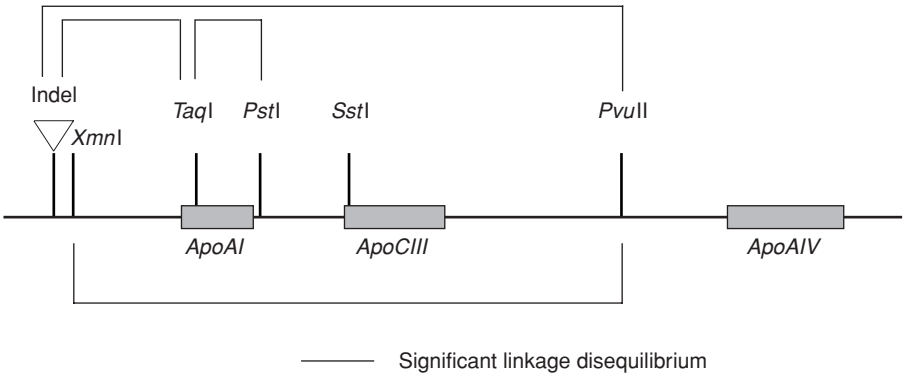

Figure 2.7. Linkage disequilibrium in autosomal *ApoAI, ApoCIII, ApoAIV* gene region of human genome between insertion/deletion polymorphism (indel) and five restriction sites. Data from Haviland et al. (1991).

and the *Pvu*II site. Hence, linkage disequilibrium in this DNA region is not a reliable guide to physical proximity.

The example shown in Figure 2.7 illustrates that evolutionary history can obscure information; in this case, we cannot infer relative positional information from the magnitude of linkage disequilibrium. This has important implications in medical genetics. Many studies will look at a few markers within a gene region and then look for disease associations. The rationale for such studies is not that the markers being scored are actually causing the disease; rather, the hope is that one or more of the markers will be in linkage disequilibrium with the mutation that actually causes the disease or contributes to disease risk (this topic will be discussed in more detail in Chapter 10). This approach has found many associations between markers and diseases in humans and illustrates just one practical application of linkage disequilibrium. However, the causative mutation is not necessarily physically closest to the marker showing the strongest disease association. As a hypothetical example, suppose the three rather evenly spaced restriction site markers *Xmn*I, *Sst*I, and *Pvu*II were used to look for disease associations between the *ApoAI, CIII, AIV* region and coronary artery disease (such associations do indeed exist). Suppose further that the insertion/deletion polymorphism was a causative mutation by affecting, say, the expression of *ApoAI*. However, using these three markers, the effect of the insertion/deletion polymorphism would be detected as an association through linkage disequilibrium with the *Pvu*II site—the site that is actually the most distant site from the hypothetical causative mutation among all the markers surveyed! Hence, an investigator might be tempted to conclude that the causative site is close to the *ApoAIV* locus and is not in the *ApoAI* locus. Ignoring evolutionary history can lead to many false conclusions in medical genetics. Our understanding of the present must be predicated upon a knowledge of how the current genetic variation arose during its evolutionary past. When dealing with a DNA region in which there is a poor or no correlation between linkage disequilibrium and physical position, we must always be cautious in interpreting marker association data in genetic disease studies.

The linkage disequilibrium found in small DNA regions can, if used properly, actually help us in our search for disease associations. For example, loci coding for apoproteins have been associated not only with coronary artery disease but also with Alzheimer's dis-

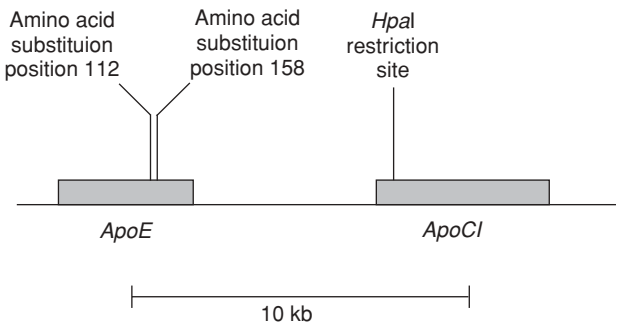

Figure 2.8. Genetically variable sites in autosomal *ApoE, ApoCI* region of human genome that have been associated with risk for Alzheimer's disease.

ease, a mental dementia that afflicts many people as they age. Figure 2.8 shows another autosomal apoprotein coding region in the human genome, this segment coding for apoprotein E and apoprotein CI. As shown, there are two polymorphic sites in the coding region of the *ApoE* locus, each associated with an amino acid change. There is extensive linkage disequilibrium between these two sites such that only three of the four possible gamete types defined by these two polymorphisms exist and $|D'| = 1$ (indeed, there is no evidence that recombination has ever occurred between these polymorphic nucleotide sites). These three gamete types define three distinct alleles at the *ApoE* locus named *ε2, ε3*, and *ε4*. Many studies have revealed an association between the *ε4* allele and a high risk for Alzheimer's disease (such an association does *not* mean that *ε4* causes Alzheimer's disease). Nearby, a *Hpa*I restriction site polymorphism was found in the *ApoCI* locus that also has a significant association with risk for Alzheimer's disease (Chartier-Harlin et al. 1994), with people bearing the allele in which the enzyme *Hpa*I cuts this site having increased risk. Note that both of these single-locus studies subdivide people into two risk categories (high and low) for Alzheimer's disease. However, the *Hpa*I site is also in linkage disequilibrium with the *ApoE* sites, so does this *ApoCI* site actually provide new risk information or is it redundant with the information previously documented in *ApoE*? When the *ApoE* alleles and the *ApoCI* restriction site are considered simultaneously, three, not two, risk categories for Alzheimer's disease are revealed (Table 2.4) (Templeton 1995). So when combined, these two apoprotein loci do indeed provide more refined information about risk for Alzheimer's disease than either locus separately. Moreover, note from Table 2.4 that the *Hpa*I restriction site is associated with both the highest and the lowest risk categories, depending upon which *ApoE* allele it is combined with. On the basis of the associations of the *Hpa*I restriction site alone, we would have placed into the "high" *Hpa*I risk category both those people

Table 2.4. Risk for Alzheimer's Disease as Associated with *ε2, ε3*, and *ε4* Alleles at *ApoE* Locus and *Hpa*I Restriction Site Polymorphism in *ApoCI* Locus

Risk Category	*ApoE*	*ApoCI*
High	*ε3* or *ε4*	*Hpa*I cuts
Medium	*ε3*	*Hpa*I does not cut
Low	*ε2*	*Hpa*I cuts

Note: See Figure 2.8.

with *highest* and *lowest* risk to Alzheimer's disease. Thus, ignoring evolutionary history as manifested in linkage disequilibrium could have led to erroneous medical advice in this case.

As shown in this chapter, the Hardy–Weinberg law, a seemingly simple model, nevertheless leads to many important insights about the evolutionary process. This model played an important role in the establishment of Mendelian genetics and natural selection during the first half of the twentieth century. The two-locus version of this law is currently playing a critical role in medical genetics in the twenty-first century. The difference in the potential for evolutionary change between the one-locus and two-locus versions shows that we must be cautious in generalizing inferences from our reductionist models. It is therefore critical to examine what happens when some of the other assumptions of the original Hardy–Weinberg model are altered or relaxed. In the next chapter, we focus upon one of these critical assumptions: system of mating.

3

SYSTEMS OF MATING

As defined in Chapter 1, a deme is a collection of interbreeding individuals of the same species that live in sufficient proximity that they share a common system of mating—the rules by which pairs of gametes are chosen from the local gene pool to be united in a zygote. Sufficient proximity depends upon the geographical range of the group of individuals and their ability to disperse and interbreed across this range. These geographical factors will be dealt with in Chapter 4. Here we simply note that, depending upon the geographical scale involved and the individuals' dispersal and mating abilities, a deme may correspond to the population of the entire species or to a subpopulation restricted to a small local region within the species' range. A deme is not defined by geography but rather by a shared system of mating. The Hardy–Weinberg model assumes one particular system of mating—random mating—but many other systems of mating exist. Moreover, as shown in Chapter 2, it is possible for different loci or complexes of loci within the same deme to have different systems of mating. It is therefore more accurate to say that a deme shares a common system of mating for a particular genetic system or locus. The purpose of this chapter is to investigate some alternatives to random mating and their evolutionary consequences.

INBREEDING

In its most basic sense, **inbreeding** is mating between biological relatives. Two individuals are related if among the ancestors of the first individual are one or more ancestors of the second individual. Because of shared common ancestors, the two individuals could share genes at a locus that are identical copies of a single ancestral gene (via premise 1—DNA can replicate). Such identical copies due to shared ancestry are said to be **identical by descent**. In contrast, the same allele can arise more than once due to recurrent mutation. Identical

Population Genetics and Microevolutionary Theory, By Alan R. Templeton

copies of a gene due to recurrent mutation from different ancestral genes are said to be **identical by state**.

Virtually all individuals within most species are related to all other individuals if you go far enough back in time. For example, computer simulations using reasonable assumptions about humanity's demographic history indicate that all humans living today share at least one common ancestor who lived sometime between 55 CE (Common Era) and 1415 BCE (Before the Common Era) (Rohde et al. 2004). Thus, all humans are biological relatives if we could trace our ancestry back a few thousand years. In practice, we often know pedigree relationships only for a few generations into the past. Given our ignorance about long-term pedigrees, how do we decide who is a relative and who is not? The solution to this practical problem is to regard some particular generation or set of individuals as the reference population whose members are regarded as unrelated. We assume that we can ascertain the biological relatedness of any two individuals in the current population by going back to but not beyond the individuals in that reference population. By assumption, all the genes in this reference population are regarded as *not* being identical by descent. If two identical genes today are traced back to different genes in the reference population, this pair of genes is regarded as being identical by state and not by descent.

There are several alternative ways of measuring inbreeding within this basic concept of mating between known relatives. Many of these alternatives are incompatible with one another because they focus on measuring different biological phenomena that are associated with matings between relatives. Unfortunately, all of these alternative ways of measuring "inbreeding" are typically called "inbreeding coefficients" in the population genetic literature. This lack of verbal distinctions between different biological concepts has resulted in confusion and misunderstanding. Jacquard (1975) tried to clarify this confusion in an excellent article entitled "Inbreeding: One Word, Several Meanings," but the many meanings of the word *inbreeding coefficient* are still rarely specified in much of the population genetic literature. The responsibility for making the distinctions among the several distinct and mutually incompatible inbreeding coefficients therefore often falls upon the reader. Consequently, it is important to be knowledgeable of the more common concepts of inbreeding, which we will examine in this chapter.

Definitions of Inbreeding

Pedigree Inbreeding. When two biological relatives mate, the resulting offspring could be homozygous for an allele through identity by descent. In other words, the gene at a particular autosomal locus being passed on by the father could be identical to the homologous gene being passed on by the mother because both genes are identical copies of a single piece of DNA found in a common ancestor. The amount of inbreeding in this case is measured by F (the first of many inbreeding coefficients), defined as the probability that the offspring is homozygous due to identity by descent at a randomly chosen autosomal locus. Offspring for whom $F > 0$ (that is, offspring with a finite chance of being homozygous at a locus through identity by descent) are said to be **inbred**. Because F is a probability, it can range in value from 0 (no chance for any identity by descent) to 1 (all autosomal loci are identical by descent with certainty). The probability F can be calculated for an individual by applying Mendel's first law of 50–50 segregation to the pedigree of that individual.

As an example, consider the pedigree in Figure 3.1, which shows an offspring produced by a mating between two half sibs. For pedigree data, the reference population is simply the set of individuals for which no further pedigree information exists. In Figure 3.1, the

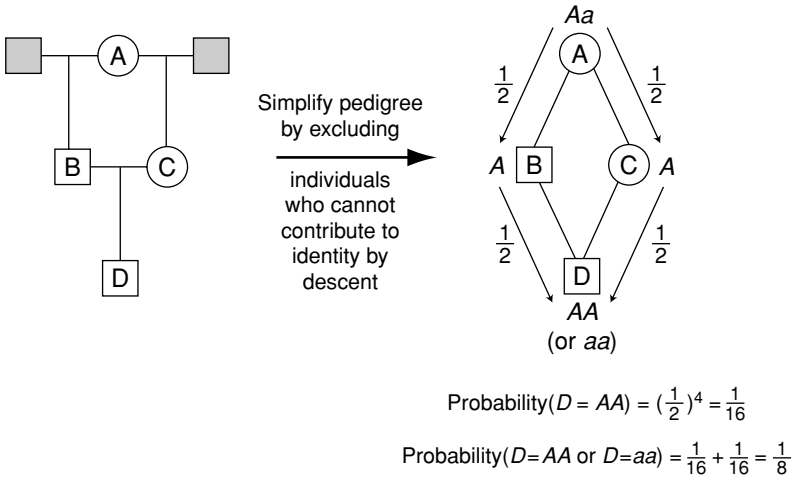

Figure 3.1. Mating between two half sibs (individuals B and C) who share a common mother (individual A, who is heterozygous *Aa*) to produce an inbred offspring (individual D). The left side of the figure portrays the pedigree in the standard format of human genetics, where squares denote males, circles denote females, horizontal lines connecting a male and female denote a mating, and vertical lines coming off from the horizontal mating lines indicate the offspring. The right side of the figure shows how this pedigree is simplified for the purposes of calculating the inbreeding coefficient F by deleting all individuals from the pedigree who are not common ancestors of the offspring of interest (individual D in this case). Shading in the pedigree on the left indicates the deleted individuals. The Mendelian probabilities associated with transmitting the *A* allele are indicated in the simplified pedigree.

reference population consists of individual A and the two males with whom she mated. These three individuals are assumed to be unrelated, and any alleles they carry, even if identical, are not considered to be identical by descent but rather to be identical by state. In Figure 3.1, there is only one shared ancestor (A) common to both the mother (C) and the father (B). Assuming that the common ancestor herself has no inbreeding in the pedigree sense, her two alleles at an autosomal locus cannot be identical by descent and are indicated by A and a (they may be identical by state). The probability that the common ancestor (A) passes on the A allele to her son (B) is $\frac{1}{2}$ from Mendel's first law, and likewise the probability that she passes on the A allele to her daughter (C) is $\frac{1}{2}$. Both the son (B) and daughter (C) also received an allele at this locus from their fathers, who are not common ancestors and cannot contribute to identity by descent. Therefore, the only way for the offspring (D) to be identical by descent for this locus is for both the father (B) and the mother (C) to pass on the allele they inherited from their common ancestor (A), and each of these gamete transmissions also has a probability of $\frac{1}{2}$ under Mendel's first law (Figure 3.1). Because the four segregation probabilities shown in Figure 3.1 are all independent, the probability that all four occurred as shown is $(\frac{1}{2})^4 = \frac{1}{16}$ = probability that individual D is homozygous by descent for allele A. The common ancestor (A) also had a second allele a, and the probability that individual D is homozygous by descent for allele a is likewise $\frac{1}{16}$. Hence, the total probability of individual D being identical by descent at this locus is $\frac{1}{16} + \frac{1}{16} = \frac{1}{8}$ since the event of D being AA is mutually exclusive from the event of D being aa. By definition, the pedigree inbreeding coefficient for individual D is therefore $F = \frac{1}{8}$.

The calculation of F can become much more difficult when there are many common ancestors and ways of being identical by descent and when the common ancestors themselves

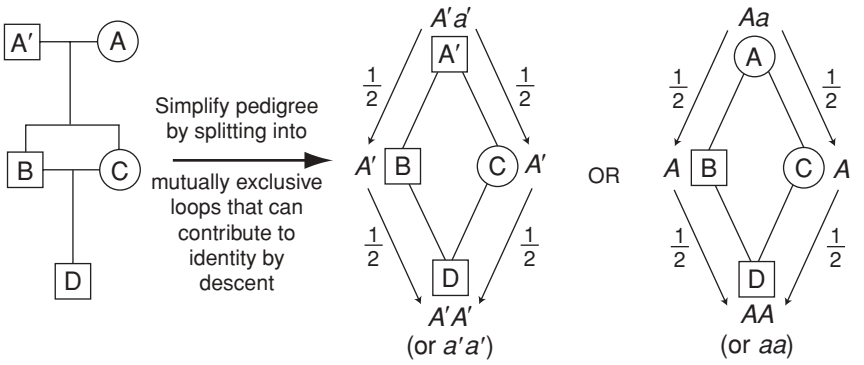

Figure 3.2. Inbreeding associated with mating of two full sibs.

are inbred in the pedigree sense. However, the basic principles are the same: Nothing more than Mendel's first law is applied to the pedigree to calculate the pedigree inbreeding coefficient F. For example, consider the case of two full sibs mating to produce an inbred offspring (Figure 3.2). In this case, the inbred offspring can be homozygous by descent for an allele from its grandmother or grandfather (Figure 3.2). Since an individual can be homozygous by descent for an allele from one and only one of the common maternal/paternal ancestors, identity by descent for an allele from the grandmother is mutually exclusive from identity by descent for an allele from the grandfather. Hence, the total probability of identity by descent, regardless of which common ancestor provided the allele, is the sum of the identity probabilities associated with the grandmother and grandfather, each of which is $\frac{1}{8}$ (Figure 3.2). Hence, $F = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$ for the offspring of two full sibs.

Of course, some pedigrees have many more common ancestors and pathways of potential identity by descent, making the calculation of F more difficult than the simple examples shown in Figures 3.1 and 3.2. The algorithms used to make these calculations for more complicated pedigrees were worked out many centuries ago by the Roman Catholic Church. Dispensations for incestuous marriages were needed to be granted before the Church could recognize such marriages. Therefore, priests needed to work out the degree of inbreeding that would occur in the offspring from such a marriage in order to distinguish degrees of consanguinity that are dispensable from those that are not (Cavalli-Sforza and Bodmer 1971). Today, many computer programs use these same algorithms to calculate F.

It is critical to note that the pedigree inbreeding coefficient F is applied to a particular *individual* coming from a specified union with a specified pedigree. Therefore F is an individual concept and not a population concept at all. Indeed, a single population often consists of individuals showing great variation in their F's. For example, a captive herd of Speke's gazelle (*Gazella spekei*) was established at the St. Louis Zoo between 1969 and 1972 from one male and three females imported from Africa (Templeton and Read 1994). Assuming that these four imported animals are unrelated (that is, the four founding animals constitute the reference population), their initial offspring would all have $F = 0$. However, because there was only one male in the original herd, the most distant relationship among captive-bred animals is that of a half sib (all the initial captive-bred offspring must share the same father). As a consequence, once the initial founders had died or were too old to breed, the *least inbred* mating possible among the captive-born animals would be between half

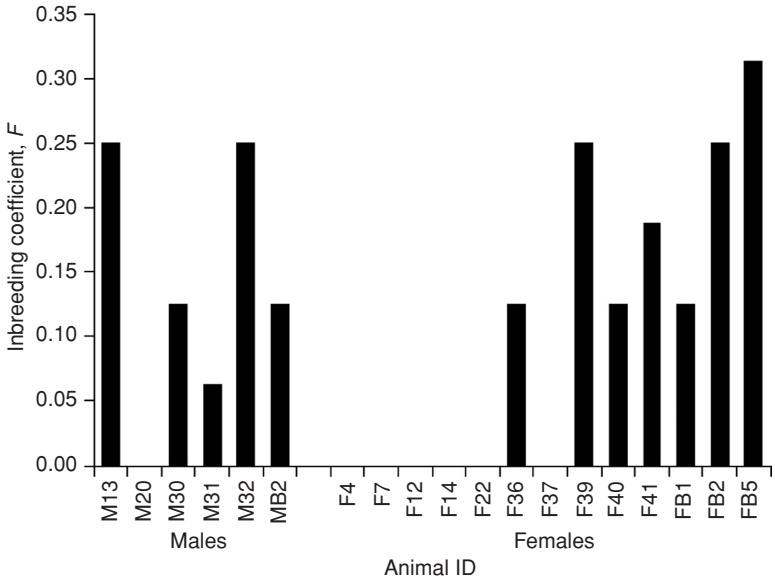

Figure 3.3. Pedigree inbreeding coefficients for all individuals from captive herd of Speke's gazelle.

sibs, with $F = \frac{1}{8} = 0.125$ (Figure 3.1). Moreover, in the initial decade of captive breeding, some father–daughter matings and other highly consanguineous matings occurred as well, resulting in a herd by 1979 (now split between zoos in St. Louis and Texas) that consisted of 19 individuals with a broad spread of individual F's ranging from 0 to 0.3125 (Figure 3.3).

Recall from Chapter 2 that the system of mating used in the Hardy–Weinberg law is a population concept applied to the level of a deme and to a particular locus. Random mating as a concept is meaningless for specific individuals within a deme. Figure 3.3 illustrates that F refers to *individuals*, not the *deme*. Hence, pedigree inbreeding (the one most people think of when they encounter the word "inbreeding") does not—indeed, cannot—measure the system of mating of a deme. This means that F cannot be used to look for deviations from the Hardy–Weinberg assumption. However, this does not mean that pedigree inbreeding has no population genetic or evolutionary implications.

One of the most important evolutionary implications of pedigree inbreeding (F) is that it displays strong interactions with rare, recessive alleles and epistatic gene complexes. Consider first a model in which a recessive allele is lethal when homozygous. Let B be the sum over all loci of the probability that a gamete drawn from the gene pool bears a recessive lethal allele at a particular locus. Because B is a sum of probabilities of non–mutually exclusive events, B can be greater than 1. Indeed, the simplest biological interpretation of B is that it is the average *number* of lethal alleles over all loci borne by a gamete in the gene pool. When pedigree inbreeding occurs, then BF is the rate of occurrence of both gametes bearing lethal alleles that are identical by descent, thereby resulting in the death of the inbred individual. Of course, an individual can die from many causes, not just due to identity by descent for a lethal allele. The only way for an individual to live is (1) not to be identical by descent for a lethal allele and (2) not to die from something else, either genetic or environmental.

Under the assumption that B is a small number, the number of times an inbred individual will be identical by descent for a lethal allele will follow a distribution known as the Poisson distribution (Appendix 2). The only way for the individual not to die of identity by descent for a lethal gene is to have exactly zero lethal genes that are identical by descent and therefore homozygous. This probability equals e^{-BF} under the Poisson distribution. Let $-A$ be the natural logarithm of the probability of not dying from any cause other than being homozgyous for a lethal recessive allele that is identical by descent. Then, the probability of not dying from something else is e^{-A}. To be alive, both events must be true, so the probability of being alive is $e^{-BF}e^{-A} = e^{-A-BF}$. Therefore, we have the expected mathematical relationship

$$\ln(\text{probability of an inbred individual with } F \text{ being alive}) = -A - BF \tag{3.1}$$

Note that equation 3.1 predicts that viability (the probability of being alive at a given age) should decrease with increasing inbreeding (as measured by F). This is an example of **inbreeding depression**, the reduction of a beneficial trait (such as viability or birth weight) with increasing levels of pedigree inbreeding. Inbreeding depression does not always occur with pedigree inbreeding, nor is it necessarily associated with any of the other definitions of inbreeding. However, inbreeding depression is a common phenomenon in mammals (Ralls et al. 1988), including humans, so we need to examine the application of equation 3.1.

One complication of applying equation 3.1 is that any one individual is either dead or alive, so the realized probability for any one individual is either 0 or 1 regardless of F. However, equation 3.1 predicts a linear relationship between the natural logarithm of the probability of being alive with F, so in the model this probability can take on intermediate values between 0 and 1. Although F is defined for an individual, equation 3.1 cannot be meaningfully applied to an individual. We must therefore extend the concept of pedigree inbreeding up to the level of a deme before we can make use of equation 3.1.

To illustrate how to do this, consider the 1979 population of Speke's gazelle whose individual F's are portrayed in Figure 3.3. As can be seen, several animals have identical levels of pedigree inbreeding: Seven animals share an $F = 0$, five share an $F = 0.125$, and four share an $F = 0.25$. Although any one animal is either dead or alive at a given age, the proportion of animals alive at a given age in a cohort that shares a common level of pedigree inbreeding varies between 0 (everyone in the cohort is dead) and 1 (everyone in the cohort is alive). Hence, the probability of an inbred individual with a specific F being alive at a given age is estimated by the proportion of the cohort that all share that same specific F that are alive at the given age. Complications can arise due to small sample sizes within certain cohorts, but small sample size corrections can be used to deal with these difficulties (Templeton and Read 1998). Equation 3.1 is now implemented by doing a regression (Appendix 2) of the natural logarithm of the cohort viability at a given age against the various F's associated with different cohorts to estimate A and B. For example, for the Speke's gazelle herd up to 1982, a regression of the natural logarithm of survivorship up to 30 days after birth upon F yields $A = 0.23$ and $B = 2.62$, and survivorship up to one year (the approximate age of sexual maturity in this species) yields $A = 0.42$ and $B = 3.75$ (Figure 3.4). This means that the average gamete from this population behaved as if it bore 3.75 alleles that would kill before one year of age any animal homozygous for such an allele.

In general, lethality can arise from several other genetic causes under inbreeding besides homozygosity for a recessive, lethal allele. For example, homozygosity for an allele may lower viability but may not necessarily be absolutely lethal. Nevertheless, homozygosity for such deleterious alleles could reduce the average survivorship for a cohort of animals

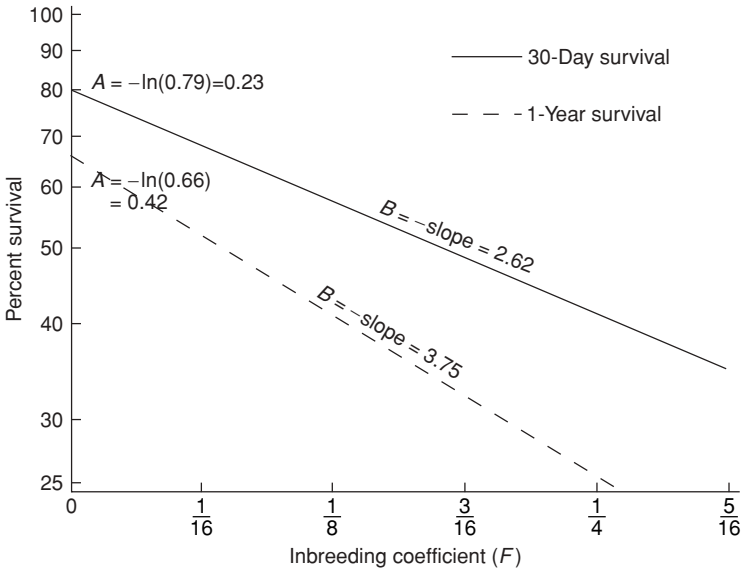

Figure 3.4. Inbreeding depression in captive herd of Speke's gazelle. From Templeton and Read (1984).

sharing a common F. Alternatively, some homozygous combinations of alleles at different loci may interact to reduce viability through epistasis. For example, knockout (complete loss of function) mutations were induced for virtually all of the 6200 genes in the yeast (*Saccharomyces cerevisiae*) genome. Yeast can exist in a haploid phase that genetically mimics the state of $F = 1$ for every locus, so the effects of these knockout mutants could be studied in the equivalent of a homozygous state. Given the compact nature of the yeast genome, it was anticipated that most of these knockouts would have lethal consequences; that is they would behave as recessive, lethal alleles. Surprisingly, more than 80% of these knockout mutations were not lethal and seemed "nonessential" (Tong et al. 2001). However, when yeast strains were constructed that bore pairs of mutants from this nonessential class, extensive lethality emerged from their interactions (Tong et al., 2001, 2004). Similarly, a detailed analysis of the genetic causes of the inbreeding depression found in the captive population of Speke's gazelle revealed that epistasis between loci was a significant contributor to the observed B (Templeton and Read, 1984; Templeton, 2002a). The yeast experiments and the results obtained with the Speke's gazelle make it clear that B should be regarded as the number of "lethal equivalents" rather than the number of actual lethal alleles. The term *lethal equivalents* emphasizes that we really do not know the genetic architecture underlying inbreeding depression from these regression analyses, but lethal equivalents do allow us to measure the severity of inbreeding depression in a variety of populations using the standard reference model of equation 3.1.

Because each diploid animal results from the union of two gametes and, by definition, the only animals that survive are those not homozygous for any lethal equivalent, a living animal is expected to bear about $2B$ lethal equivalents in heterozygous condition. In the original non-inbred population of Speke's gazelles, the average number of lethal equivalents for one-year survivorship borne by the founding animals of this herd is therefore 7.5 lethal equivalents per animal. Studies on inbreeding in humans from the United States and Europe yield values of $2B$ between 5 and 8 (Stine 1977). These numbers mean that most humans, just like most Speke's gazelles, are bearers of multiple potentially lethal genetic diseases

or gene combinations. Consequently, there is a large potential for inbreeding depression and other deleterious genetic effects in most human populations when pedigree inbreeding does occur. For example, cousin matings represent only 0.05% of matings in the United States (Neel et al. 1949), but 18–24% of albinos and 27–53% of Tay-Sachs cases (a lethal genetic disease) in the United States come from cousin matings (both of these are autosomal recessive traits, with the recessive allele being rare). This same pattern is true for many other recessive genetic diseases. Hence, even small amounts of pedigree inbreeding in a population that is either randomly mating or even avoiding system-of-mating inbreeding (to be discussed next) can increase the incidence of some types of genetic disease by orders of magnitude in the pedigree-inbred subset of the population.

Because of inbreeding depression and the tendency for increased incidence of genetic disease in consanguineous matings in humans, "inbreeding"—regardless of the exact definition being used—is often viewed as something deleterious for a population. The idea that inbreeding is deleterious has raised many concerns for endangered species, as such species often are reduced to small sizes, which as we have seen leads to pedigree inbreeding. Studies on pedigree inbreeding depression, such as those performed for Speke's gazelle, demonstrate that these concerns are real, and much of applied conservation genetics focuses on dealing with inbreeding in its various senses and consequences. However, is inbreeding always deleterious? The answer appears to be no. For example, many higher plants have extensive self-mating, the most extreme form of inbreeding, and this inbreeding can be adaptive under many conditions (Holsinger 1991). To understand the ultimate cause for why inbreeding is not always deleterious, we must turn our attention from inbreeding at the level of an individual to inbreeding at the level of a deme's system of mating.

Inbreeding as Deviation from Random-Mating Expectations. To obtain a system of mating measure of inbreeding at the deme level, we must examine deviations from Hardy–Weinberg genotype frequencies that are due to nonrandom mating. First, recall the random-mating model for the simple one-locus, two-allele (A and a) model shown in Table 2.2. Note that in Table 2.2 the genotype frequencies are obtained by multiplying the allele frequencies associated with the male and female gametes. Now, suppose that gametes are put together in such a way that there is a deviation from the product rule of Hardy–Weinberg in producing genotype frequencies but that the marginal allele frequencies remain the same. Let λ be this deviation parameter from the simple product of the gamete frequencies, as shown in Table 3.1. Note from Table 3.1 that λ only affects the genotype frequencies and not the gamete frequencies. This is because λ is designed to measure how gametes come together to form genotypes for a given set of gamete frequencies. Also, note that λ is applied to a deme and measures deviations from Hardy–Weinberg genotype frequencies in that deme. In contrast, F is defined for individuals, not demes, and measures the probability of identity by descent for that individual and not the system of mating of the deme as a whole. Biologically, λ is quite different from F.

Mathematically, λ is also quite different from F. Recall that F is a probability and like all probabilities is defined only between 0 and 1 inclusively. In contrast, as can be seen in Box 3.1, λ is the covariance (see Appendix 2) between uniting gametes. A covariance is proportional to the correlation coefficient (Appendix 2) and can take on both positive and negative values and is mathematically noncomparable to a probability such as F. If $\lambda > 0$, there is a positive correlation between uniting gametes in excess of random-mating expectations. This means that the alleles borne by the uniting gametes are more likely to share the same allelic state than expected under random mating. If $\lambda < 0$, there is a negative correlation between uniting gametes, and the alleles borne by the uniting gametes are less

Table 3.1. Multiplication of Allele Frequencies Coupled with Deviation from Resulting Products as Measured by 1 to Yield Zygotic Genotypic Frequencies under System of Mating That Allows Deviation from Random Mating.

			Male gametes		
	Allele		A	a	
	Frequency		p	q	
	Allele Frequency				Marginal allele frequencies in deme
Female gametes	A	p	AA $p^2 + \lambda$	Aa $pq - \lambda$	$(p^2 + \lambda) + (pq - \lambda)$ $= p^2 + pq = p(p + q)$ $= p$
	a	q	aA $qp - \lambda$	aa $q^2 + \lambda$	$(qp - \lambda) + (q^2 - \lambda)$ $= qp + q^2 = q(p + q)$ $= q$
Marginal allele frequencies in deme			$(p^2 + \lambda) + (qp - \lambda)$ $= p^2 + qp = p(p + q)$ $= p$	$(pq - \lambda) + (q^2 + \lambda)$ $= pq + q^2 = q(p + q)$ $= q$	

Summed frequencies in zygotes:

AA: $G'_{AA} = p^2 + \lambda$
Aa: $G'_{Aa} = pq - \lambda + qp - \lambda = 2pq - 2\lambda$
aa: $G'_{aa} = q^2 + \lambda$

Note: The zygotic genotype frequencies are indicated by G'_k.

likely to share the same allelic state than expected under random mating. Random mating occurs when there is no correlation between uniting gametes ($\lambda = 0$). The actual correlation between uniting gametes is $\lambda/(pq)$ (see Box 3.1). The correlation coefficient (Appendix 2) has a standardized range of -1 to $+1$ inclusively, in contrast to the covariance that has no standardized range. Hence, it is more convenient to measure deviations from Hardy–Weinberg at the deme level in terms of the correlation of uniting gametes as opposed to the covariance of uniting gametes. Accordingly, we define the inbreeding coefficient to be $f \equiv \lambda/(pq)$, defined as the correlation of uniting gametes within the deme. From Table 3.1, we can now see that the genotype frequencies that emerge from this system of mating can be expressed as

$$\begin{aligned} G'_{AA} &= p^2 + \lambda = p^2 + pq\left(\frac{\lambda}{pq}\right) = p^2 + pqf \\ G'_{Aa} &= 2pq - 2\lambda = 2pq - 2pq\left(\frac{\lambda}{pq}\right) = 2pq - 2pqf = 2pq(1 - f) \\ G'_{aa} &= q^2 + \lambda = q^2 + pq\left(\frac{\lambda}{pq}\right) = q^2 + pqf \end{aligned} \tag{3.2}$$

Because f is a correlation coefficient, it can take on both positive and negative values (as well as zero, the random-mating case). Generally, when f is positive, the system of mating of the deme is described as one of inbreeding, and when f is negative, the system of mating of the deme is described as one of avoidance of inbreeding. However, regardless of whether or not f is positive or negative, f is called the inbreeding coefficient.

BOX 3.1 THE CORRELATION OF UNITING GAMETES

In order to show that λ is the covariance among uniting gametes, we must first define a random variable to assign to the gametes. In our simple genetic model, the gametes bear only one of two possible alleles, A and a. Let x be a random variable that indicates the allele borne by a male gamete such that $x = 1$ if the male gamete bears an A allele and $x = 0$ if the male gamete bears an a allele. Similarly, let y be a random variable that indicates the allele borne by a female gamete such that $y = 1$ if the female gamete bears an A allele and $y = 0$ if the female gamete bears an a allele. Let p be the frequency of $A-$bearing gametes in the gene pool. Because we are dealing with an autosomal locus, p is the frequency of A for both male and female gametes.

Using these definitions and the standard formula for means, variances, and covariances (Appendix 2), we have

$$\begin{aligned}
\text{Mean}(x) &= \mu_x = 1 \times p + 0 \times q = p \\
\text{Mean}(y) &= \mu_y = 1 \times p + 0 \times q = p \\
\text{Variance}(x) &= \sigma_x^2 = (1-\mu_x)^2 \times p + (0-\mu_x)^2 \times q \\
&= (1-p)^2\, p + (-p)^2\, q = pq \\
\text{Variance}(y) &= \sigma_y^2 = pq \\
\text{Covariance}(x, y) &= (1-\mu_x)(1-\mu_y)(p^2+\lambda) + (1-\mu_x)(0-\mu_y)(2pq-2\lambda) \\
&\quad +(0-\mu_x)(0-\mu_y)(q^2+\lambda) \\
&= q^2(p^2+\lambda) - pq(2pq-2\lambda) + p^2(q^2+\lambda) \\
&= \lambda(q^2+2pq+p^2) \\
&= \lambda
\end{aligned}$$

Hence, λ is the covariance between uniting gametes under a system of mating that produces the genotype frequencies given in Table 3.1. Because covariances do not have a standardized range whereas correlations do, it is usually more convenient to measure the nonrandom associations between uniting gametes through their correlation coefficient rather than their covariance. The correlation coefficient is (Appendix 2)

$$\rho_{x,y} = \frac{Cov\,(x, y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{\lambda}{pq}$$

Although inbreeding as measured by f alters the genotype frequencies from Hardy–Weinberg (equations 3.2), it does not cause any change in allele frequency. The frequency of the A allele in the final generation in Table 3.1 is

$$\begin{aligned}
p' &= 1 \times (p^2 + pqf) + {}^1/_2[2pq(1-f)] = p^2 + pqf + pq(1-f) \\
&= p^2 + pqf + pq - pqf = p^2 + pq = p(p+q) = p
\end{aligned}$$

Because the allele frequencies are not changing over time in Table 3.1, inbreeding as measured by f is not an evolutionary force by itself at the single-locus level (that is, system-of-mating inbreeding alone does not change the frequencies of alleles in the gene pool).

Another interpretation of f is suggested by equations 3.2: In addition to f being a correlation coefficient, f is also a direct measure of the deviation of heterozygote genotype frequencies from Hardy–Weinberg expectations. Note that the frequency of heterozygotes in equations 3.2 is $2pq(1-f)$, and recall from Chapter 2 that the expected frequency of heterozygotes under Hardy–Weinberg is $2pq$. Hence, an alternative mathematical definition of f is

$$f = 1 - \frac{\text{observed frequency of heterozygotes in deme}}{\text{expected frequency of heterozygotes under Hardy–Weinberg}} \tag{3.3}$$

From equation 3.3, we can see that a positive correlation between uniting gametes leads to a heterozygote deficiency in the deme (typically called an inbreeding system of mating), no correlation yields Hardy–Weinberg frequencies (random mating), and a negative correlation (typically called avoidance of inbreeding) yields an excess of heterozygotes in the deme.

In most of the population genetic literature, both f and F are called inbreeding coefficients and are often assigned the same mathematical symbol. That will not be the case in this book. The symbol F, which will be called pedigree inbreeding, refers to a specific individual, measures that individual's probability of identity by descent for a randomly chosen autosomal locus, and ranges from 0 to 1. In contrast, f will be called system-of-mating inbreeding, refers to a deme, measures deviations from Hardy–Weinberg genotype frequencies, and ranges from -1 to $+1$ (Table 3.2). Because f and F are both called inbreeding coefficients and frequently assigned the same symbol in much of the literature, it is not surprising that these two extremely different definitions of inbreeding have often been confused. We will illustrate the difference between these two inbreeding coefficients by returning to the example of the captive herd of Speke's gazelle.

Recall that the captive herd of the Speke's gazelle was founded at the St. Louis Zoo with one male and three females between 1969 and 1972. Because there was only one male, all animals born in this herd were biological relatives. Under the assumption that the four founding animals were unrelated (our reference population), all of these original founders and the offspring between them have $F = 0$; that is, these individuals were not "inbred." By 1982, these older animals had all died off and all animals in the herd had $F > 0$. Given

Table 3.2. Contrast between Pedigree Inbreeding Coefficient *F* and System-of-Mating Inbreeding Coefficient *f*

Property	F	f
Data used to calculate	Pedigree data for specific individuals	Genotype frequency data for specific locus and deme
Type of mathematical measure	Probability	Correlation coefficient
Range of values	$0 \leq F \leq 1$	$-1 \leq f \leq 1$
Biological level of applicability	Individual	Deme
Biological meaning	Expected chance of identity by descent at randomly chosen autosomal locus for specific individual caused by biological relatedness of individual's parents	System of mating of deme measured as deviations from random-mating genotype frequency expectations

that all animals bred in captivity had to be at least half sibs of one another (there was only one founding male), this inbred state of the descendants of the original founders and their offspring was inevitable *regardless* of system of mating. The average F in 1982 was 0.149 relative to the founder reference population, making this captive herd one of the most highly inbred populations of large mammals known. An isozyme survey (Appendix 1) was also performed on these same animals in 1982. For example, at the polymorphic *general protein* locus (*GP*), the observed heterozygosity was 0.500, but the expected heterozygosity under random mating was 0.375. Hence, for this locus, $f = -0.333$. Several other polymorphic isozyme loci were scored, all yielding $f < 0$, with the average f over all loci being -0.291. This highly negative f indicates a strong avoidance of system-of-mating inbreeding.

We now have what appears to be a contradiction, at least for those who confuse f and F. This herd of gazelles is simultaneously one of the most highly inbred (pedigree sense F) populations of large mammals known and is also strongly avoiding inbreeding (system-of-mating sense f). There is no paradox here except verbally; the two types of inbreeding and inbreeding coefficients are measuring completely different biological attributes. The negative f indicates that the breeders of this managed herd were avoiding inbreeding in a system-of-mating sense within the severe constraints of this herd of close biological relatives.

If inbreeding were being avoided at the level of system of mating, then why did every individual in the herd have an $F > 0$? Keep in mind that "random mating" means that females and males are paired together "at random" regardless of their biological relationship. In any finite population, there is always a finite probability of two related individuals being paired as mates under random mating. The smaller the population, the more likely it is to have biological relatives mate at random. Hence, random mating ($f = 0$) implies some matings among biological relatives that will yield $F > 0$ in any finite population. Indeed, even avoidance of inbreeding ($f < 0$) can still result in matings among biological relatives in a finite population. For example, many human cultures (but not all) have incest taboos that often extend up to first cousins. Assuming a stable sized population of N adults with an average and variance of two offspring per family (the number of offspring being Poisson), then $f = -1/(N - 10)$ when relatives up to and including first cousins are excluded as mates but mating is otherwise random (Jacquard 1974). Note that as N increases, f approaches 0. This means that although incest taboos are common in human societies, the Hardy–Weinberg law fits very well for most loci within most large human demes. However, some human demes are small. Suppose $N = 50$ (a small local human population, but still found in some hunter/gathering societies); then $f = -0.025$ under this nonrandom system of mating. Nevertheless, such small human populations typically contain many inbred ($F > 0$) individuals despite their incest taboos in choosing mates ($f < 0$).

Consider, for example, a set of religious colonies in the upper great plains of North America that are descendants of a small group of anabaptist Protestants who originally immigrated from the Tyrolean Alps (Steinberg et al. 1966). There has been very little immigration into these religious colonies from other human populations, so they represent a genetic isolate. Internally, their system of mating is one of strong avoidance of mating between close relatives as incest is considered a sin. Despite this strong avoidance of pedigree inbreeding, the average F for one isolated subsect was 0.0255. This makes this population one of the more highly inbred human populations known despite a system of mating that avoids inbreeding. The reason for this seeming contradiction is that these colonies were founded by relatively few individuals, so virtually everyone in the colony today is related to everyone else. Hence, the pedigree inbreeding is due to the small population size at

the time the colonies were founded and not due to the system of mating (such "founder effects" will be discussed in more detail in the next chapter). Indeed, if these individuals truly mated at random, then the average F under random mating would be 0.0311, a value considerably larger than the observed average F of 0.0255. As this population reveals, avoidance of inbreeding in the system-of-mating sense does *not* necessarily result in no pedigree inbreeding ($F = 0$) but rather results in lower levels of pedigree inbreeding than would have occurred under random mating. The strong avoidance of inbreeding in this human population also results in large deviations from Hardy–Weinberg expectations. For example, a sample from this population scored for the MN blood group had 1083 individuals with genotype *MM*, 1220 with *MN*, and 260 with *NN*. Using the test given in Chapter 2, the resulting chi square is 9.68 with one degree of freedom, which is significant at the 0.002 level. Hence, unlike most other human populations (see Chapter 2 for two examples), this religious colony does not have Hardy–Weinberg genotype frequencies for the *MN* locus. Instead, there is a significant excess of heterozygotes (only 1149 are expected under random mating, versus the 1220 that were observed). Using equation 3.2, this results in $f = -0.0615$. Thus, this religious colony started from a small number of founders is highly inbred in the pedigree sense ($F = 0.0255$), even though the population is strongly avoiding inbreeding in the system-of-mating sense ($f = -0.0615$). The two inbreeding coefficients F and f are most definitely not the same either mathematically or biologically in this human population.

Another human example is given in Figure 3.5 (Roberts 1967) that illustrates how small founding population size can result in pedigree inbreeding despite strong avoidance of system-of-mating inbreeding. Twenty people colonized the remote Atlantic island of Tristan da Cunha in the early 1800s, with a few more migrants coming later (more details will be given in the next chapter). Despite a strong incest taboo among these Christian colonists and a system of mating characterized by $f < 0$, individuals with pedigree inbreeding began to

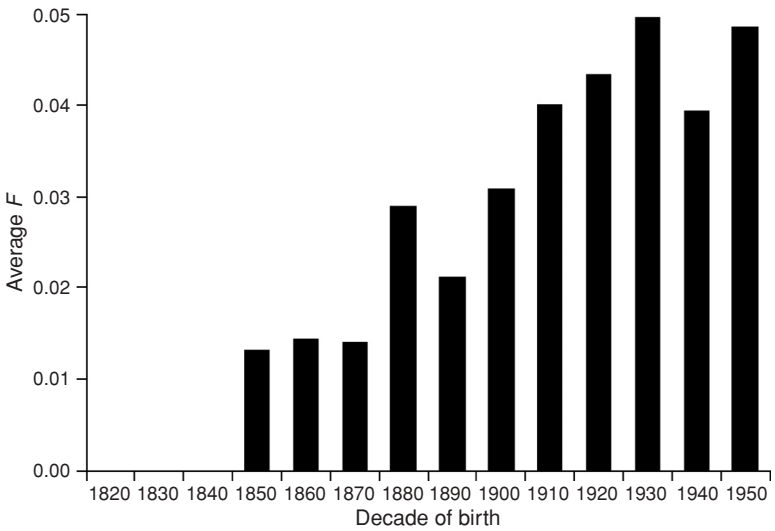

Figure 3.5. Average pedigree inbreeding coefficient for human population on Tristan da Cunha as function of decade of birth.