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

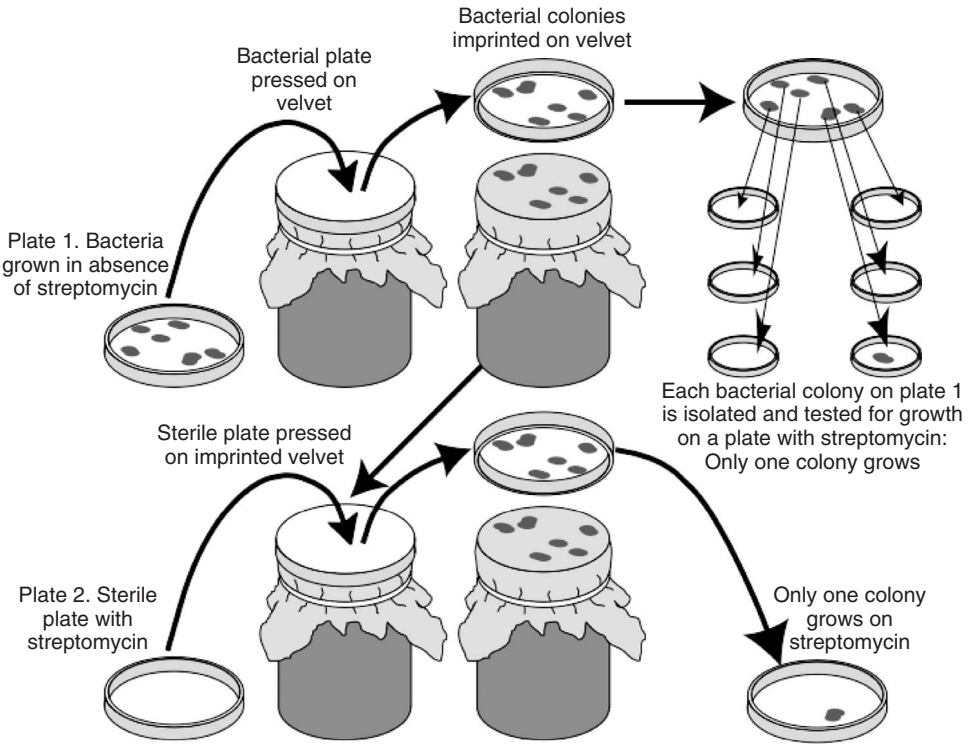

Figure 1.1. Replica plating. A suspension of bacterial cells is spread upon a Petri dish (plate 1) such that each individual bacterium should be well separated from all others. Each bacterium then grows into a colony of genetically identical individuals. Next, a circular block covered with velvet is 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

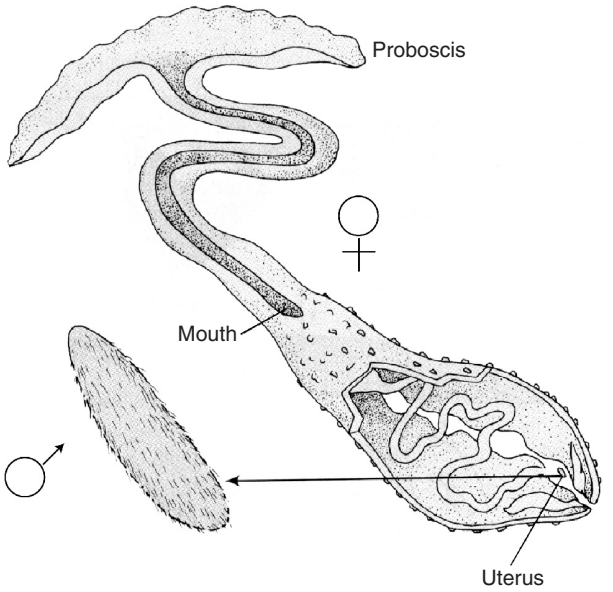

Figure 1.2. Sexes in *Bonellia*. The female has a walnut-sized body that is usually buried in the mud with a protruding proboscis. The male is a ciliated microorganism that lives inside the female. Adapted from Fig. 3.18 from *Genetics*, 3rd Edition, by Peter J. Russell. Copyright © 1992 by Peter J.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

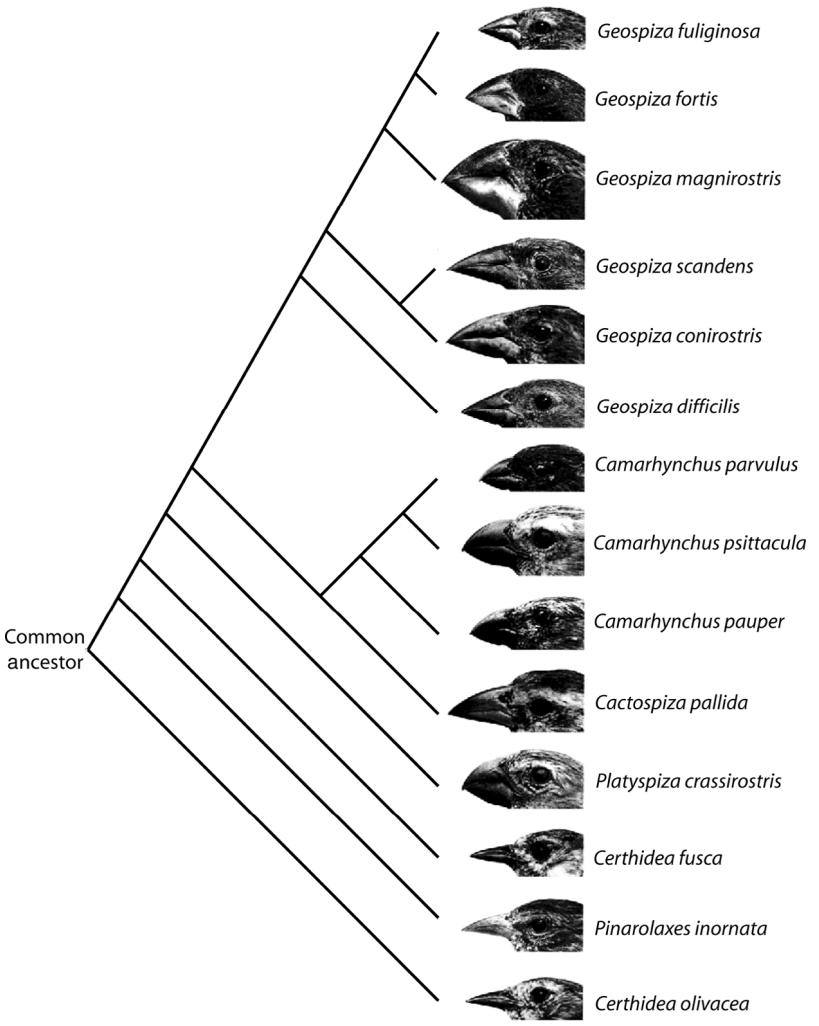

Figure 1.3. Evolutionary tree of 14 species of Darwin's finches estimated from molecular genetic data. Modified from Fig. 3 in Petren et al. (1999). Copyright ©1999 by the Royal Society of London.

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

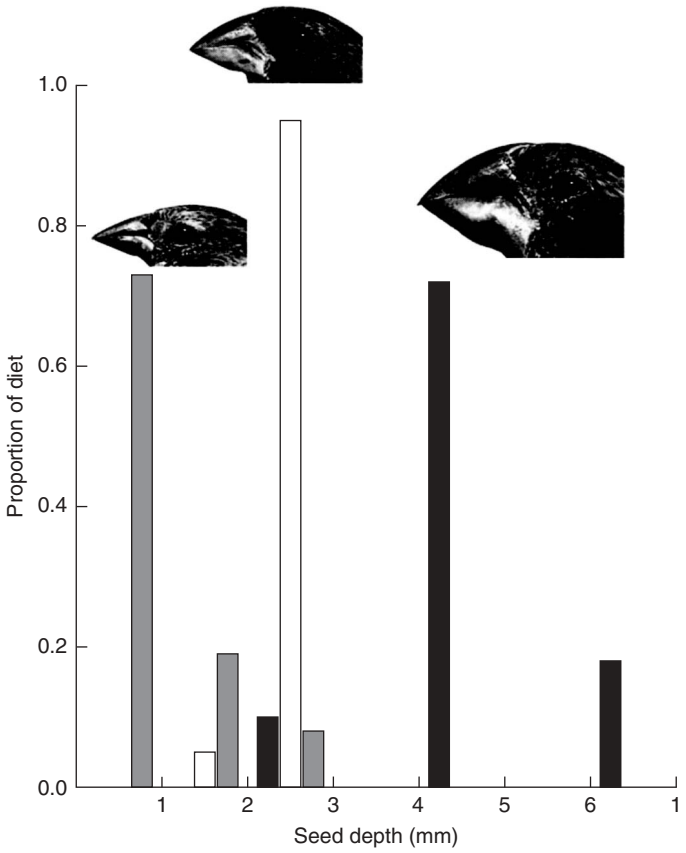

Figure 1.4. Proportions of various seed sizes in diet of three of Darwin's finches: *Geospiza magnirostris* (solid bars), *Geospiza fortis* (open bars), and *Geospiza fuliginosa* (gray bars). Redrawn with permission from Fig. 35 in P. R. Grant, *Ecology and Evolution of Darwin's Finches* (1986). Copyright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

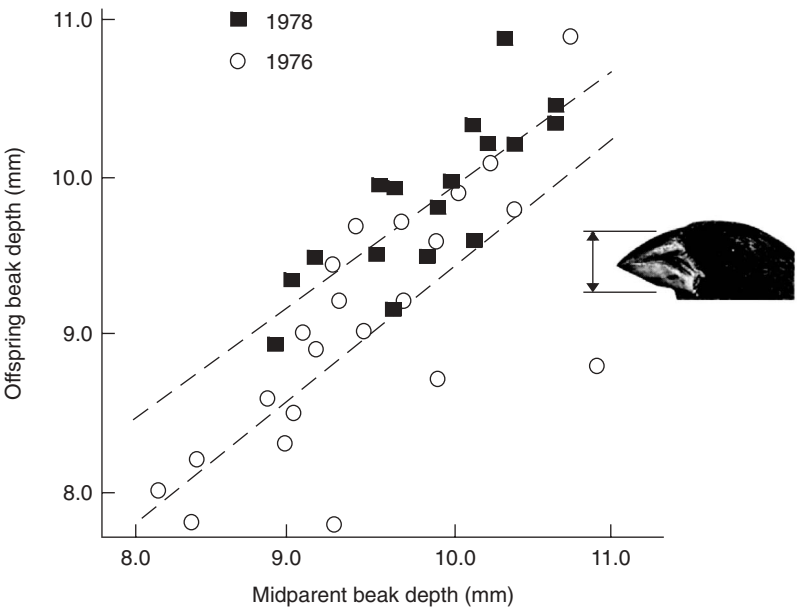

Figure 1.5. Relationship between beak depth of offspring and average beak depth of their parents (midparent beak depth) in medium ground finch, *G. fortis*, as measured in two years, 1976 and 1978. The lines show a fitted least-squares regression to these data (Appendix 2). As will be explained in Chapter 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

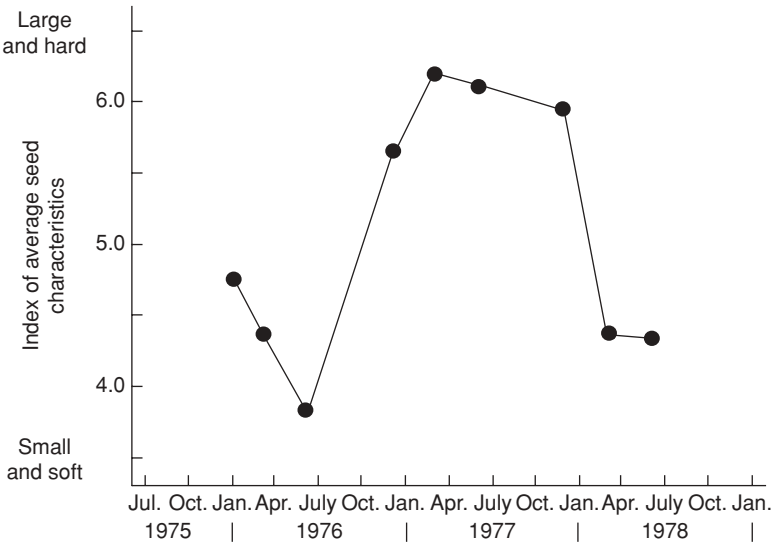

Figure 1.6. Characteristics of average seed available as food to medium ground finches (*G. fortis*) before, during, and after 1977 drought. Reprinted from Fig. 1 in P. T. Boag and P. R. Grant, *Science* 214: 82–85(1981). Copyright ©1981 by the AAAS.

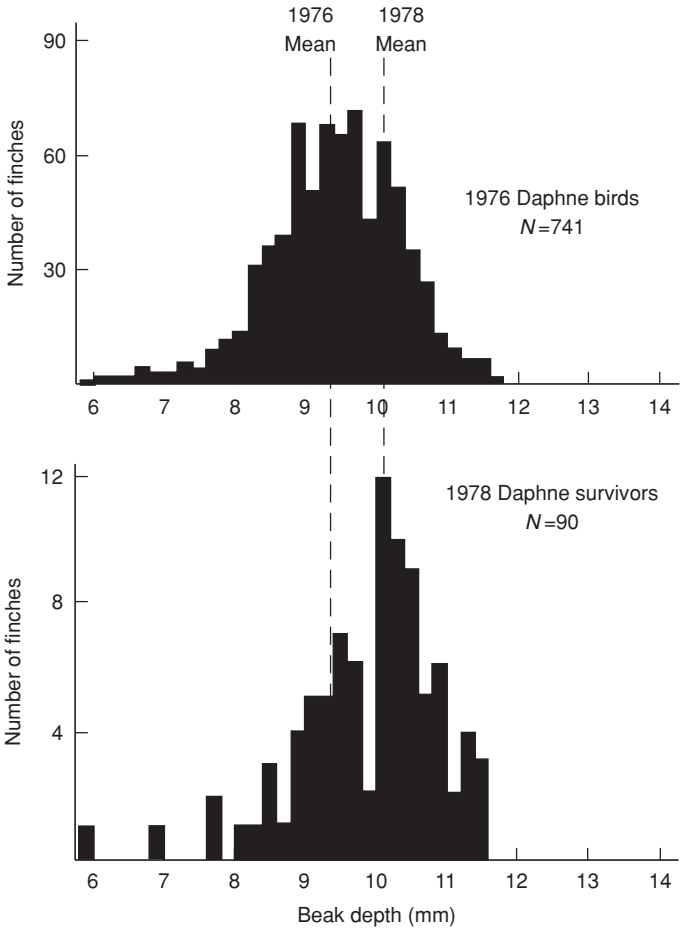

Figure 1.7. Frequency distributions of beak depth in *G. fortis* on island of Daphne Major before (1976) and after (1978) a drought. Dashed lines indicate the mean beak depths in 1976 and 1978. Redrawn with permission from Fig. 59 in P. R. Grant, *Ecology and Evolution of Darwin's Finches* (1986). Copyright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

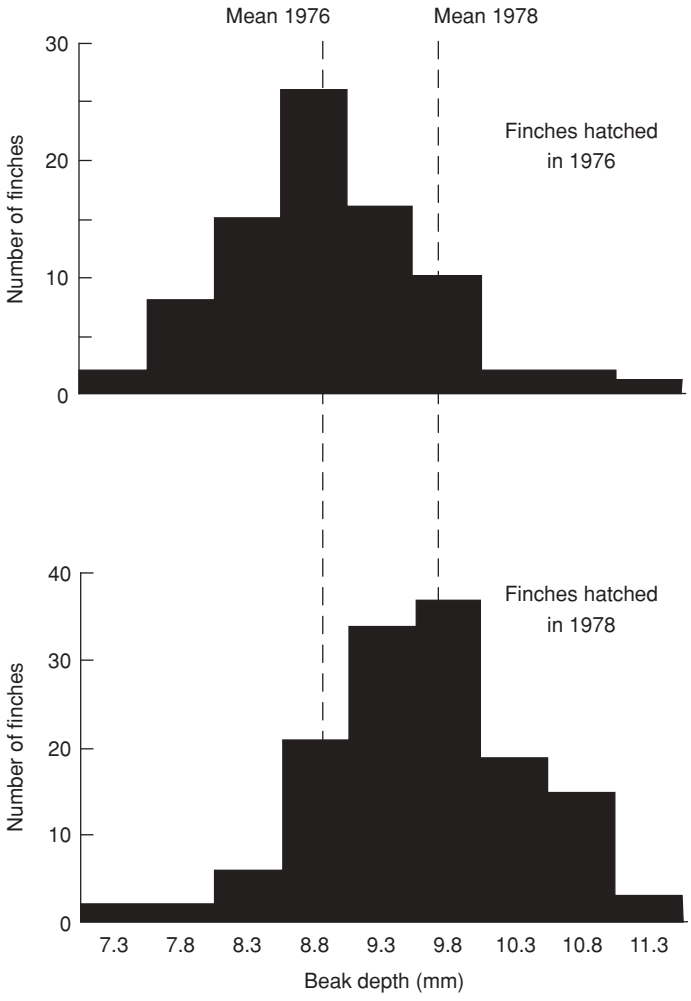

Figure 1.8. Beak depth in *G. fortis* hatched year before drought (1976) versus year after drought (1978). Dashed lines indicate the mean beak depth for the finches born before and after the drought.

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.