

# PREFACE

## WHY POPULATION GENETICS, WHY NOW?

This Fourth Edition of *Principles of Population Genetics* provides an introduction to population genetics as it exists today. It is intended for students at the undergraduate and graduate levels as well as newcomers from fields like human genetics. The presentation is balanced between theory and observation, and has been written to optimize its use as a teaching aid. Our emphasis is on critical and analytical thinking about real data. We have therefore included numerous problems interspersed within the chapters to encourage students to stop for a moment to think about what they have just learned and how it applies to real situations.

Modern population genetics has been revamped and revitalized by three different but related revolutions in ideas and technology. One revolution was *conceptual* and turned theoretical population genetics on its head. Another was *empirical* and gave population geneticists vast amounts of data to analyze. And the third, *computational*, allowed the conceptual advances and the new data to come together.

The conceptual revolution emerged from the theory of coalescence, which established a framework for thinking about populations and genes in terms of their evolutionary history. This seemingly simple change in perspective was of major importance. Because each of the alleles present in a population at any time descends from an allele that existed in the past, the composition of a sample of alleles taken from a population at any time contains important



information about the evolutionary history of the alleles, and therefore information about the history of the population.

The empirical revolution resulted from genomics, which changed population genetics in three distinct but critically important ways. The first was the determination of the complete DNA sequences of many genomes, including the human genome and that of key model organisms along with those of closely related species. The second was the use of DNA sequencing technology to discover single-nucleotide polymorphisms (SNPs, pronounced *snips*) throughout the genome. Such polymorphisms are important because one of the goals of population genetics is to understand the evolutionary and biological significance of genetic variation. The third contribution of genomics was its high throughput and cost-effective methods for determining the genotypes of hundreds of thousands of SNPs in hundreds or thousands of individuals. This technology provides population geneticists with the substrate for identifying genes affecting complex traits in humans and other organisms. These traits include genetic risk factors for disease, genes affecting polygenic traits in human parasites or agriculturally important plants and animals, and genes that have been targets of natural selection, including those genes whose activities distinguish the physical and mental development and behavior of humans from our closest primate relatives.

The third revolution was computational, including not only vastly increased power for sheer number-crunching but also the explosion of the Internet as a medium for communication among researchers as well as a repository of data and other informational resources. The increased computational power has resulted in widespread application of such methods as Markov-chain Monte Carlo simulation, which allows maximum likelihood and Bayesian methods to be applied to huge data sets for statistical inference on a scale that not long ago would have been considered unimaginable. As one insightful high-school senior has described statistics, "All unrigged data is messy in its own right, and statistics is a collection of methods by which to gather and make sense of that mess artfully in at least one helpful and meaningful way." To make optimal use of statistical resources on the Internet, the text includes references to many vital web resources that connect the material to up-to-the-minute progress in modern population genetics.

- Chapter 1 examines the fundamental observations that motivated the origins of population genetics. It includes the apparent disconnect between dichotomous phenotypic variation of the sort studied by Mendel and continuous phenotypic variation of the sort that so fascinated Galton. The chapter discusses various hypotheses for the origin and maintenance of genetic variation as well as the ways that modern molecular methods are used to discover and quantify the amounts and structure of genetic variation in natural populations.
- Chapter 2 focuses on the relationship between allele frequencies and genotype frequencies, with special emphasis on random mating. We consider statistical pitfalls of testing for random mating both when small



sample sizes and when large-scale simultaneous tests are carried out on hundreds of thousands of SNPs throughout the genome. The chapter also discusses linkage disequilibrium, its causes and measurement, why it matters, and how it is used.

- Chapter 3 considers random genetic drift in the context of the classical Wright-Fisher model. We discuss the mechanisms that cause random genetic drift and examine how diffusion theory was originally deployed to obtain key theoretical results. We then introduce the concept of coalescence and show how this approach is particularly apt for the analysis of molecular data.
- Chapter 4 adds mutation and lays down the foundations for the neutral theory of molecular evolution. We show how mutation and random genetic drift act together to determine the fate of neutral or nearly neutral alleles, and we also consider the characteristics of samples drawn from natural populations at steady state.
- Chapter 5 addresses the mechanisms of Darwinian selection in detail. The chapter brings together the theoretical foundations and empirical observations of the dynamics of gene-frequency change under the action of natural selection. We develop haploid and diploid cases and the concepts of equilibrium, stability, and context dependence. Not only do we develop classical models of mutation-selection balance, but we also consider models of natural selection in many other guises.
- Chapter 6 examines population subdivision and its consequences for the distribution of genetic variation among subpopulations, including the hierarchical  $F$  statistics used in estimating these effects. We examine migration in the context of population subdivision, along with the most up-to-date methods for estimating migration rates from natural populations.
- Chapter 7 focuses on molecular evolutionary genetics, currently one of the most active areas for theory and experiment. The unifying theme is the neutral theory; we closely examine the correspondence between the data and theory. Our goal is to give the student a firm grasp of the fundamentals and a deep enough understanding of the principles to identify important gaps in our knowledge.
- Chapter 8 covers the implications of quantitative genetics from an evolutionary perspective. A compelling argument for using quantitative genetics for the study of evolution is that adaptive evolution takes place at the level of the phenotype, and quantitative genetics provides the tools for understanding transmission of phenotypic traits. We give theoretical quantitative genetics special importance because of the paradoxes it raises in contrasting evolution at the levels of the phenotype and of the DNA sequence.
- Chapter 9 is completely new. It covers the exciting field of population genomics, which refers to the analysis of population genetic principles on a genome-wide scale. Included are discussions of genome size and com-



position, genome-wide patterns of polymorphism, linkage disequilibrium, evolutionary divergence, various aspects of the population genetics of speciation, and an extended discussion of the population genetics of transposable elements.

- Chapter 10 is also completely new and shows how the principles of population genetics are actively being applied to human populations to identify genetic risk factors for complex diseases and to understand human evolutionary history and migration across the planet.

Throughout the book we have taken special care to explain the biological motivation behind the theoretical models so that the models do not simply materialize out of thin air, and to explain in plain English the implications of the results. Many concepts are illustrated by numerical examples, using actual data wherever possible. Special topics and examples are often set off from the text as boxed problems whose solutions are explained step by step. Every chapter ends with about 20 problems, graded in difficulty, and solutions worked in full appear at the end of the text.

## ACKNOWLEDGMENTS

DLH is deeply indebted to four long-term collaborators—Elena R. Lozovsky, Stanley A. Sawyer, Dyann F. Wirth, and Sarah K. Volkman—as well as to numerous graduate students and postdoctoral fellows, past and present, for their continuing support, as well as for sharing their infectious energy, curiosity, creativity, imagination, and the fruits of their hard work. They are too numerous to name individually, but please consult [www.oeb.harvard.edu/hartl/lab/](http://www.oeb.harvard.edu/hartl/lab/) for a roster as well as brief descriptions of our current research projects. Special thanks to current postdoctoral fellows and especially to current graduate students Trevor Bedford, Kyle Brown, Mauricio Cameiro, Marna Costanzo, Kerry Geiler, Matthew Hegreness, Sarah Kingan, Christian Landry, Bernardo Lemos, Rebekah Rogers, and Martine Zilversmit for their patience and understanding.

AGC would like to thank his current lab members, especially graduate students Erin Hill, Amanda Larracuenta, Kirk Lohmueller, Tim Sackton, Sarah Stockwell, and Xu Wang for their inspiring drive, independence, hard work, and their understanding during a particularly challenging rewrite. Other members of my lab group, past and present, have been a great source of motivation to make this book reflect the joy and depth of pursuing a life in population genetics research. Please see [www.mbg.cornell.edu/faculty-staff/faculty/clark.cfm](http://www.mbg.cornell.edu/faculty-staff/faculty/clark.cfm) for details.

The authors also gratefully acknowledge the professionalism and patience of the Sinauer team, especially Kathaleen Emerson and Sydney Carroll for overtime work and attention to detail, Janice Holabird for clean and uncluttered pages, Chris Small for production, Bobbie Lewis for copyediting, and Andy Sinauer for encouragement and support.