Introduction

Genomes and genomic variation entered into the study of genetic variation in natural populations in this century. The human genome sequencing projects led to increasingly affordable procedures for studying sequence variation, and by 2001[1] population genetic studies of genes were already dominated by analyses of sequence variation. During the work on the human sequence more than a million places in the genome were discovered to vary among 24 individuals representing the ethnic variation in the world.[2] This corresponds to one single nucleotide polymorphism for every 2000 base pairs in the DNA of the human genome, corresponding to a recombination distance of less than one crossover in 10,000 meioses—quite a dense map, and even denser ones are now available because more genomes are sequenced, and hence more single nucleotide polymorphisms are revealed as differences among more people.[3] The genome sequence of the common chimpanzee has also been determined, allowing the recent evolution of the human genome sequence to be addressed. Scores of genomes in other animals, fungi, and plants have been sequenced, and even more are currently under way.

Such a brief account can only provide a superficial impression of the amount of population data currently available, presently accumulating at an immense rate, and expected to keep increasing in the foreseeable future. This offers ample opportunity for investigations into the history and dynamics of gene and genome variation in natural populations. Fundamental questions of evolution may be asked and new ones formulated. To profit from this wealth of data, however, great challenges have to be overcome. Population genetics, as most branches of genetics, is thus in the most exciting of circumstances for a scientific field. For scientists the wish "may you live in interesting times" is surely not a curse—given, of course, that scientific matters catch the public awareness.

Population genetics has been around for about a century, emerging right at the dawn of genetics. The field can be defined as the study of the distribution of hereditary variation across time and space in species and populations. A human population is in this context a biologically reasonable assemblage of people. Individuals usually find their mates within their own population, and it commonly comprises all humans inhabiting a more or less well-defined area.

[1] The draft human genome was published by Lander et al. (2001) and Venter et al. (2001).
[2] Sachidanandam et al. (2001).
[3] Every person adds new genetic variation (Levy et al. 2007)

All of humanity may in some circumstances be considered a population, but usually a more restricted definition is used, for instance a city and its environs, an island or peninsula, or a continent. Denmark, Faroe Islands, and Greenland are three distinct populations in the Kingdom of Denmark, even though each is of mixed origin and regularly receives immigrants from the others, and from the rest of the world, for that matter. Each of the three may be further separated into local populations on islands—real islands in Denmark and the Faroes, and islands of habitable areas in Greenland.

The description of population variation has two foci: the general understanding of biological evolution and the application of genetic variation for human welfare. Mendelian genetics supplied Darwin's theory of evolution with crucial elements, many of which were developed and matured by population geneticists in the decades before the so-called neo-Darwinian synthesis around 1930. Medical applications are centered on the understanding of the prevalence of rare genetic and hereditary diseases, and on the basis of the hereditary aspects of many common diseases.[4] This has driven recent developments in human genetics, not least the sequencing of the human genome. Genetic investigations of human diseases necessarily require population genetic studies—inheritance and segregation are observed in existing families. The methods used in these areas have benefited the development of other aspects of population genetics, in particular applications to animal and plant breeding.

The field of population genetics builds on experiences from observations and experiments and is supported by a well-developed framework of theory founded early in the twentieth century.[5] The basic laws of genetic transmission are probability laws. Genetics and, in particular, population genetics has thus always relied on observations interpreted through statistical analysis, and many developments in statistics have their origin in genetic applications. Population genetic theory entertains the whole spectrum characteristic of population sciences from statistical modeling to descriptive dynamic theory. Through time, the weight of the statistical and descriptive aspects of theory have changed—as well as the interest in theoretical or empirical developments of the field in general. Surprisingly, these two historical oscillations seem largely uncorrelated, but at present we are in a statistical and empirical era due to important breakthroughs within both of these approaches in the recent past.

The analysis and interpretation of data on molecular genetic variation relies on population genetic theory, coalescent theory in particular, and the implementation and execution of such analysis requires skills in computer science and statistics. Much of this activity occurs within the field of bioinformatics. Accordingly, the writing of this volume was largely carried out at the Bioinformatics Research Center (BiRC) at the University of Aarhus. One aim is to communicate some of the experiences from a century of population genetics, and relate them to contemporary developments for the benefit of students and

[4]Everyone is expected to carry genetic determinants of rare diseases and susceptility genes for common diseases. This expectation is typified in genome sequences of the individual (Levy et al. 2007)

[5]See Provine (1971).

colleagues in bioinformatics. A second aim is to participate in the process of incorporating some of the recent developments of population genetics into biology teaching.

The two aims seem contradictory, and the only reason I attempt to combine them is that my students taught me that it is possible. For some years, students of bioinformatics have attended the biology course, which developed into the present book, and biology students frequented courses in the elements of bioinformatics. Population genetics is a field of study where formal and quantitative theory play an integral part. The language of such theory includes mathematical formulations and reasoning, and throughout its history the field has attracted attention from students with a background in mathematics and its applications. Population genetic theory, however, resides firmly within biology, because its issues arise from biological phenomena, and its results refer to those phenomena. Any biology student of population genetics considers theory as a background for his or her activities—variation exists only in the level of theory deemed necessary.

The structure of the book tries to accommodate the broad range of potential readers. Short introductions to genetic subjects and concepts are given at appropriate places to avoid the overwhelming task of studying an introductory text in genetics. In a course setting, students with a biological background naturally offer assistance—and welcome the recap of basic genetics. The mathematical requirements correspond to the introductory mathematic courses given in undergraduate biology teaching. They include few formal requirements beyond elementary algebra and calculus, but assume that fear of adding and multiplying letters has been alleviated. Students of biology, many of whom are decidedly not theoretically inclined, have used the ancestral lecture notes with success, but they needed to realize that the biological content is the focus. The level of statistical background seems to vary considerably more among biology students than does their baground in mathematics, so the necessary statistical concepts are briefly explained in Appendix A. Still, with a reasonable consensus on minimal requirements, a textbook should meet and challenge the students on their home ground, while keeping the material palatable for all readers. In the running text calculations and mathematical arguments are relegated to text boxes, and if their contents require difficult or lengthy arguments, they are marked by a $*$ in front of the box title. Most are, however, fairly uncomplicated and do not require a high level of mathematical abilities—many of the arguments may look more complicated for biologists than they really are. Their presence is intended to tickle the curiosity of readers with a more mathematical background. On a similar note, some of the scattered exercises and footnotes have the warning star. The answers to the exercises should often vary with the background of the student. Solutions suggested in Appendix B tend to be short and incomplete, to leave room for developments along lines of personal interests.

The focus of the book is on the theoretical background of contemporary population genetics, while acknowledging that population genetics is a subject that grew and continues to grow in the close interplay between empiricism and theory. Theoretical results refer to the material world, and observations rarely make sense without reference to the theoretical background of the field. This interplay is acknowledged in the discussion of empirical investigations that range the history of population genetics. The coverage of observations and experiments is in no way intended to be broad or representative of the field. Rather, the empirical references are chosen mainly for their qualities as illustrations of the development of thoughts within the science of population genetics.