Preface

Genetic research is at a historical crossroads. Large quantities of genetic data are being generated at an ever-increasing pace and are being used and interpreted by students and professionals from a range of disciplines, including biology, medicine, archaeology, informatics, and forensic science. Working in genetics today requires a broad skillset, from the lab bench to the computer terminal. Most importantly, though, a thorough understanding of evolutionary principles is essential for making sense of genetic data. Accordingly, evolutionary genetics is becoming a key subject for the life sciences and beyond.

DNA is an essential molecule, occurring in every living organism on this planet. It orchestrates a multitude of complex biochemical processes and pathways that ultimately result in life, allowing organisms to interact with each other in similarly complex ways. DNA is a self-replicating molecule. Every time a cell divides, its DNA is copied, base for base, with great precision, but not absolute perfection. Any copy of DNA in an organism living today is the descendent of an unbroken chain of copies ranging back to the dawn of life on Earth. A range of deterministic and stochastic processes has molded this remarkable molecule, and thus the organisms it encodes, for nearly four billion years. The anomalies, the replication errors or *mutations*, represent the very raw material for organic evolution. Processes, such as natural selection and genetic drift determine the fate of the novel varieties that arise from mutations. Over time these processes cause organisms to change and diverge from each other, giving rise to all the endless varieties of species that have occupied our planet, including the present-day biota and ourselves.

Thinking in terms of evolutionary genetics involves abstractions that challenge our intuition. First, evolutionary biology is a scientific discipline with a historical dimension. One must be able to conceptualize dynamic processes that operate at very different time scales: From instantaneous germ-line mutations and recombination events, through processes of selection and drift that alter the genetic make-up of a population over hundreds and thousands of generations, via processes of lineage splitting and speciation where the time scale may be in the orders of hundreds of thousands, or millions of years, all the way to large scale patterns of evolutionary change that span the history of life on Earth. Second, evolutionary processes involve several different levels of organization. From changes at the molecular level, to the effects these have on biochemical pathways, cells, organs, individual organisms, populations, and species.

Our ambition with this book is to help you make these abstractions and to ultimately make sense of the diverse and rich field of evolutionary genetics. We want you to understand fundamentally important processes such as mutation, natural selection, genetic drift, and speciation, and their consequences. Therefore, in the first half of the book we have focused on explaining a limited set of evolutionary models, but in some depth. Our aim is to motivate you to move beyond just understanding concepts, and to introduce you to how such concepts are applied. So, in the second half of the book, we focus more and more on how high-throughput sequencing data is changing the way evolutionary biology research is carried out. We are well aware that understanding evolutionary concepts and analysis is best done with a practical approach. Practical experience is second-to-none when it comes to developing an understanding of how to use genetic data to analyze and address interesting questions in

the life sciences and how to interpret results in meaningful ways. So, along with the main text we provide study questions to motivate you to think and reflect on the concepts in each chapter. Furthermore, we provide a series of online, computer-based practical sessions where you will be able to reinforce your understanding and gain hands-on experience with analyzing and interpreting genetic data with an evolutionary perspective. Crucially, the practical computing assignments are centered on the R programming language. This might seem a little daunting at first but it is now an essential skillset for any evolutionary geneticist. It is a tool we use every day in our own research and one we cannot stress the importance of enough.

The book and associated teaching material is based on an undergraduate course at the University of Oslo, Norway. We assume that the readers have some general knowledge in biology and genetics. Moreover, we assume some knowledge and experience in algebra, probability theory, and statistics. Inexperienced readers should not worry; we have aimed to build arguments from basic principles throughout the book.