III

Limits to Historical Inference and Prediction

The study of evolution is, in part, the study of biological history. The evolutionary geneticist has an advantage over the student of human cultural history in three major respects. First, evolutionists can manipulate their system and in effect rerun history as they try to understand the underlying mechanisms that govern change. The second advantage arises because the rules of genetic transmission are regular and susceptible to mathematical analysis. It is possible to make mathematically precise predictions about gene transmission under various sets of assumptions. The third advantage lies in the ability to sort out the temporal sequence of mutational change, which provides a crude means of dating temporal events and of ordering organismal relationships within biological groups. This record is always with us because it is part of our genetic endowment; we do not have to work from the fragments of records that have survived eons of time as do the students of human history, although ultimately the genetic record also erodes, owing to the relentless accumulation of mutations.

The mathematical biologist Gary Churchill (Chapter 6) explores the problem of inferring an ancestral character state on a evolutionary tree as a specific case of the reconstruction of history from extant information. Churchill employs a Bayesian framework to analyze several cases that include a molecular clock assumption or that dispense with a clock assumption and that feature star phylogenies versus treelike phylogenies. Churchill ends with the observation that what we conclude from such an analysis "depends on what we believed to begin with." That is not to say that we cannot illuminate history, but rather there is a kind of relativistic limitation that includes our subjective choice of model and prior distribution, and these color the final conclusions.

Ultimately, our ability to infer the past is absolutely dependent on a genetic sample drawn from extant organisms (with the minor exception of ancient DNA), and Aquadro (Chapter 7) reminds us that sampling strategy imposes a practical limitation on knowledge. We cannot know about

events not recorded in the sample. Aquadro also reminds us that mutation is a very slow phenomenon and the window of time resolved is limited by the slow accumulation of mutational change. Of course, as Aquadro notes, this problem may be mitigated by increasing sample size or by selecting DNA regions that evolve rapidly (e.g., microsatellite sequences). We also can anticipate major advances in our ability to acquire data, and this should reduce limitations associated with sample size.

Aquadro goes on to ask an important question: Have new phenomena been uncovered through research in molecular population genetics? He concludes that a new phenomenon resulting from the complex relationship between linkage and selection has emerged from molecular-level investigations. Specifically, the level of genetic diversity appears to be positively correlated with local recombination rate, and this is true for many genes and several different species. One may argue that this phenomenon was foreshadowed in theoretical work by Hill and Robertson and others (e.g., see Chapter 5, this volume), but it is important to acknowledge that there are phenomena waiting to be discovered, and present knowledge is certain to expand in ways that we may not anticipate.

Li and colleagues (Chapter 8) present a powerful example of comparative genomics in their analysis of primate color vision. Primate color vision is determined by a multigene family that includes an autosomal blue pigment gene and X-linked genes for red and green opsins. There are many tools available to the student of comparative genomics, and these are well illustrated in the chapter. The comparative approach is shown to be very powerful, because it exploits the accumulation of mutational change over extended periods of evolutionary time, and we learn, for example, that trichromancy has arisen independently by several different mechanisms in primate evolution. One particularly interesting finding is that gene conversion involving different gene family members has been important in the evolution of color vision genes, and in one case this may have produced a novel adaptive haplotype. But as Li and colleagues note, gene conversion also erases evolutionary history and renders certain details of the past unknowable. The authors also point out that there are fundamental problems with the measurement of biophysical parameters in this system that may limit precise knowledge about the way specific amino acid changes alter spectral sensitivity maxima.

The converse of historical inference is future prediction. The principles of evolutionary genetics are frequently used for prediction when the goals are utilitarian. For example, it is common for plant and animal breeders to estimate narrow heritibilities to predict short-term selection response. In recent years, students of phenotypic evolution also have applied these methods to the longer-term analysis of morphological evolution (see

Chapter 12, this volume). An area where prediction is especially important is conservation biology, where the goal is to develop strategies that will minimize threats to species survival based on the application of genetic, ecological, and demographic principles. Nunney (Chapter 9) explores this theme by focusing on effective population size (N_e) as a predictor of population survival. There are many problems with the estimation of effective population size that create uncertainties for prediction. Because the future is unknown, the conservation biologist must use uncertain estimates of the N_e parameter to develop broad-based strategies that are robust to future uncertainty and to parametric uncertainty. This dictates general rather than specific guidelines in the development of conservation strategies.

Weir (Chapter 10) considers the uses of population genetic models for the estimation of human relationships and for the estimation of linkage between disease and marker genes. He makes the strong point that many applications in this area of inference confuse single-point estimates with mean values. The reliability of inferences derived from this subtle error is unknown, although it may be possible to use computer simulation to gain some insight into the question of reliability. Weir also reminds us of the fact that all estimates of relationship are relative to some assumed base population value. This relativity may be a fundamental limitation to knowledge because the true base population is lost in history and genealogical information is beyond the reach of science.