provided, that is, that you are prepared to camp with some itinerant Australian sheep shearers. I am not sure, though, that I want to recommend such a pastime. Dr Ryder reminds us further of the lament of the shearer's wife:

> Friday night he's too tired; Saturday night too drunk;
> Sunday too far away.

Perhaps that is why he called his book 'Sheep & Man', though there may be another reason on pp. 315–316, which I must allow the reader to look up for himself. After all, and unlike Dr Ryder, I am not prepared to go into good black print and slur the name of a smart Highland regiment. Good grief, no.

R. C. ROBERTS
Chief Scientists' Group
Ministry of Agriculture, Fisheries & Food

Statistical Analysis of DNA Sequence Data. Edited by B. S. Weir, Marcel Dekker, AG Verlag/Publishers, Elisabethenstrasse 19, Postfach 133, 4010 Basel, Switzerland, 1983. Pp. 255 Price $45.00. ISBN 0 8247 7032 3

The recent advances in molecular genetical techniques have already resulted in large quantities of detailed information on genomic structure. Restriction enzyme map data and sequence data are accumulating rapidly: in DNA sequence data alone, over two million bases are stored in the EMBL nucleotide sequence library (version 4, August 1984). The rate of increase is staggering: DNA sequence data is growing at the rate of 0·8 million bases per year.

For workers in molecular biology and molecular evolution this is a time of immense opportunity. The interpretation of such data requires not only new conceptual frameworks but also new statistical techniques. The analysis of the new molecular genetic data also reveals many problems of interest to applied statisticians. It is to these three groups that 'The Statistical Analysis of DNA Sequence Data', edited by B. S. Weir, is meant to appeal.

Most of the authors are, however, in the field of population genetics or molecular evolution, and this is evident throughout the book. Six of the nine chapters are concerned with statistical problems of an evolutionary nature; the other three sit uncomfortably alongside. Molecular biologists will find the first chapter useful. Schaffer deals with the estimation of DNA fragment lengths from mobilities on a gel: a practical problem, although somewhat peripheral to the book's main theme. Chapter two, by Gingeras, reviews the growth of computer software for sequence analysis. The available programs reflect the needs of molecular biologists. No phylogeny programs are listed. Gingeras sketches this rapidly developing area and attempts to define its future direction. The utility of this chapter is unclear: lists of software will soon date. There is little discussion of the underlying algorithms. Researchers beginning to apply computer methods will notice that the machine dependence of various packages and the subsequent difficulties in implementation on foreign machines is not mentioned. There is, however, no discussion of the statistical aspects of sequence searches. Given the intrinsic high level of noise in DNA data, this is a surprising omission.

The need for statistical methods is not recognised by many molecular biologists. The unfortunate widespread use of parsimony methods for constructing evolutionary trees is discussed by Felsenstein. Maximum likelihood methods are developed as an alternative to this arbitrary practice.

It is clear that the 'Classical' theory of population genetics, based on gene frequencies, is inappropriate for the new classes of genetic data. The mathematical models developed

https://doi.org/10.1017/S0016672300022011 Published online by Cambridge University Press

throughout the book are based on the neutral theory. Templeton, however, examines the effect of convergent evolution on estimates of genetic distance.

Molecular evolutionists will find the contributions of Ewens, Kaplan and Brown and Clegg excellent reading. Ewens provides a lucid discussion of the theoretical problems underlying the interpretation of restriction enzyme map data, concluding that no one model is best. Hence there is no best estimator of heterozygosity. Kaplan develops various models of the process of nucleotide substitution and studies the behaviour of estimators of heterozygosity and genetic distance using simulation and data analysis. Brown and Clegg note that 'the comparison of complete DNA sequence data has provided...an unprecedented view of the elemental processes of evolutionary change'. They provide a detailed analysis of sequence variation in a sample of highly repeated genes. This excellent chapter defines the kinds of variation in DNA data and provides examples of simple tests on the nature of the mutation process. The inherent problem in DNA data, small sample size, is briefly discussed.

The last two chapters, by Bishop *et al.* and by Asmussen and Clegg, examine the use of sequence data in human genetics to allow the mapping of 'disease' genes and prenatal diagnosis of genetic defects.

Overall there is an excess of model formulation and a resultant paucity of data analysis, which is a pity given the richness of sequence information. Some sources of variation in DNA sequence data are not mentioned, e.g. errors in sequencing and in transcription. This will be obvious to users of the NIH GenBank and EMBL data bases. There is also some overlap between chapters: e.g. Ewens and Kaplan both discuss estimation of heterozygosity from restriction map data.

'The Statistical Analysis of DNA Sequence Data' is the first attempt to define this new discipline. Although somewhat overpriced, it is essential reading for statisticians and workers in molecular evolution. Its appeal to molecular biologists will be small, unfortunately, and this area must await a more relevant text.

FRANK WRIGHT
Department of Animal Genetics
University of Edinburgh

https://doi.org/10.1017/S0016672300022011 Published online by Cambridge University Press