
Book Reviews

Algorithms on Strings, Trees and Sequences by DAN GUSFIELD, Cambridge University Press 1997, ISBN 0 521 58519 8, 534+xviii pages.

The academic field of bioinformatics has matured at a remarkable speed. Less than ten years ago any monograph on the subject would have been of potential interest to every practitioner. Now a book about computer methods of sequence analysis might be a ‘button-pushing’ manual for a particular suite of programs, or a work of the most technical mathematical analysis, or might occupy one of several pieces of intellectual territory lying roughly between these two.

This book is distinctly on the computer science side of the middle of this range. It attempts to analyse all the main algorithms concerned with strings of characters, which are of use or interest to molecular biology. Analysing algorithms means showing that they are correct, estimating their relative performance (speed) when applied to various problems, and identifying useful improvements and shortcuts. So it is primarily intended for those who write programs in this field to save them from having to re-invent the wheel (or rather, to save them from having to hunt through most of a Computer Science library to find it). As the author says in his Preface, it is the book he ‘wished I had available when I began learning about string algorithms’. Since that time, of course, every research group has gained ready access to developed programs performing all the main tasks of standard sequence analysis – most recently by the use of freely available Internet tools. This is a book not for experimental molecular biologists, but for computer science students who are seeking to join the growing band of bioinformatics specialists employed in research programmes with data-handling problems which lie outside the range of standard applications.

The first few Chapters deal with the ‘exact matching’ problem: how often and where does a specific short series of characters occur in a long text? The main algorithms here (‘Boyer–Moore’ and ‘Knuth–Morris–Pratt’) were invented 20 years ago, and were lying ready to hand when the invention of

DNA sequencing began to present molecular biology with questions of this kind. Gusfield develops these in an integrated and comprehensive way and, as elsewhere in the book, presents a large number of examples as exercises for the reader.

These algorithms are most appropriate when the ‘long text’ will only be searched once or a few times. In practical molecular biology applications, the text is more usually the whole database of known sequences (or a subset), and the problem is to search it repeatedly for different patterns. Under these conditions the effort required to prepare an index of the database may be repaid by much faster subsequent searches. Methods involving ‘suffix trees’ provide efficient ways of doing this, and the second section of the book, which completes roughly its first half, develops these algorithms and a large number of specific applications. An example of current particular interest is the detection of repeated elements in sequences, since when applied to the complete genome sequences of micro-organisms this may reveal features of their functional organization and evolution. The book will probably become the standard reference for suffix tree algorithms in sequence analysis, since the material has not otherwise been so conveniently brought together.

Most ‘exact-matching’ problems arise in handling nucleic acid sequences. Protein sequence comparisons lack sensitivity unless they also take account of conservative amino-acid substitutions, and of the variable length of solvent-exposed loops which lie outside the core structure of globular proteins. One approach to this problem is to perform ‘exact’ searches with many conservatively-substituted variants on the target pattern. This approach forms the basis of the ‘BLAST’ group of high-speed protein sequence database searches. The more rigorous alternative is to use one of a family of dynamic programming algorithms, and these algorithms also commonly form the basis of methods for aligning two or many members of the same protein family.

The third and largest section of the book is devoted to these subjects; the length is appropriate since their application in various guises forms the backbone of most molecular biology computing.

The dynamic programming methods are usually referred to collectively as the ‘Needleman–Wunsch–Sellers’ algorithms. Needleman and Wunsch proposed the first protein sequence comparison method of this kind; Sellers provided the first mathematical proof of the algorithms’ correctness. This book appears to make no mention of Sellers’ work, and this is unfortunate in a computer science text. A more serious limitation for its contemporary value is that there are several theoretical topics and extensions to these algorithms which are important for practical applications and which are dealt with inadequately or not at all. Examples are the statistical significance of protein sequence alignments (this shortcoming is acknowledged in the Preface), RNA secondary-structure prediction methods of Zuker, Nussinov and others, and the ‘double-dynamic-programming’ algorithm used in sequence and structure alignment by Jones, Taylor and Thornton.

The final section of the book describes topics which are developed less systematically, either because the author believes they will not be of lasting importance, or because they do not directly involve algorithms on strings of characters. These topics include physical mapping and DNA sequence assembly, and the deduction of phylogenetic trees from sequence.

This book thoroughly fulfils its purpose of systematically surveying the string algorithms which are used in biological sequence analysis. It will provide an excellent tutorial text for computer science, and a reference resource which will be valuable for practitioners writing programs in this field. It is written in a discursive and good-humoured style which makes it very easy to read, and the large number of carefully-discussed examples should allow the reader to develop real expertise relatively painlessly.

A. F. W. COULSON
Biocomputing Research Unit
Institute of Cell and Molecular Biology
University of Edinburgh

The Genetics of Sheep. Edited by L. PIPER and A. RUBINSKY. CAB International. 1997. 611 pages. Price £95. ISBN 0 85199 200 5

Sheep are probably the most versatile of the domestic animal species. In temperate countries today they are kept mainly for the production of meat, wool and milk. Particularly in some of the world’s poorest countries, they have additional, important roles, including the provision of pelts, fertilizer and fuel, providing a four-legged form of financial investment, and using resources unsuitable for other forms of agriculture. Objective methods for the genetic improvement of sheep – based on measurement of

performance in these economically important traits – have been employed for several decades in many countries. However, the refinement of existing techniques for genetic improvement and the development of new ones, including new statistical, reproductive and molecular genetic techniques, is both scientifically challenging and potentially of great economic and practical importance. This book will be valuable for all those interested in the science of sheep genetics and the practical application of this science.

The book comprises 22 chapters spanning systematics and phylogeny, genetic aspects of domestication, the genetics of colour and of morphological traits and disorders, biochemical and molecular genetics, the genetics of disease resistance, the molecular biology and genetics of scrapie, cytogenetics, physical and linkage maps, the genetics of behaviour and reproduction, reproductive technologies, developmental genetics, genetic resources and conservation, the genetic improvement of meat, milk and wool production, and genetic nomenclature. The authors are well-known authorities in their respective fields, and the book admirably achieves its purpose of providing a comprehensive reference work which will be invaluable to students and researchers in animal genetics. (The book is the first in a series with this aim – others on pigs and cattle are in preparation.) Naturally for a book of this scope, the styles and depths of different chapters vary, but overall the book is well organised and well written. Mismatched references in one or two places, and poor reproduction of figures in one chapter, are minor irritations in what is otherwise an accomplished text.

A particular aim of the book was to review the major advances in molecular genetics applied to livestock breeding, which have taken place over the last couple of decades. The chapters devoted to this area will be welcomed by those new to this field or only peripherally involved; they provide a valuable synthesis of the story so far, and a useful starting point from which to delve into the extensive specialist literature. The chapters on the molecular biology and genetics of scrapie and modern reproductive technologies are also particularly topical. The inevitable downside is that the chapters in these fast-moving areas are likely to become outdated fairly quickly – for instance, the widely-reported birth of a viable lamb (aka ‘Dolly’) following nuclear transfer of a cell derived from an adult, only just makes it into the book by way of a footnote to the chapter co-authored by two of those responsible for this development. However, even the chapters in these rapidly-changing fields will remain, for some years to come, a useful entry point for those starting a review.

In summary, this is an excellent book for students, teachers and researchers in sheep genetics. The price tag probably means that it will appear on rather more

library shelves than office shelves, but if you work in this area you will want access to it one way or the other.

GEOFF SIMM
Genetics and Reproduction Department
Animal Biology Division
Bush Estate, Penicuik
Midlothian EH26 0QE

Extreme Environmental Change and Evolution. By ARY A. HOFFMANN and PETER A. PARSONS. Cambridge University Press. 259 + xii pages. Price £55 hard £19.95 paper. ISBN 0521 44107 2 (Hardback), 0521 44659 (Paperback).

This book is based on the premise that ‘extreme environments have an important impact on evolutionary change’, and that such environments ‘have been largely ignored in the current evolutionary literature’ (p. xi). Its aim is to present facts that will convince the reader that there are special features of extreme environments that render them of exceptional interest, over and above the truism that evolution by natural selection occurs most rapidly when a population is exposed to the challenge of a novel environment, especially when adaptation to this environment requires changes in multiple characters.

The most radical possible role for sudden environmental change is that discussed in Chapter 2. This is the idea that the variability utilized by selection is itself affected by the environment. The most extreme possibility is that of ‘directed’ or ‘adaptive’ mutations i.e. mutations which confer a fitness advantage in relation to a new environmental state, and arise specifically in response to this environment. This is of course contrary to neo-Darwinian orthodoxy, based on a long line of experiments that show that mutations arise without reference to their ability to confer adaptations. While experiments on bacteria and yeast have suggested the possibility of such directed mutations, their interpretation has been fraught with difficulty and controversy. Unfortunately, Hoffmann and Parsons provide only a superficial and out-of-date review of this topic, which is surely a core issue for establishing a special role for environmental stresses in promoting evolution. A less extreme possibility is that increased rates of mutation are simply provoked by certain kinds of environmental shocks, and they review evidence that this is sometimes the case. The implication is that such mutagenesis may be adaptive, since it allows the production of the occasional mutation that promotes survival under the new environmental conditions. Unfortunately, this ignores the fact that most of the induced mutations will be harmful, and hence such increased mutation rates are unlikely to be favoured except in asexual species, where a mutator gene can remain associated

with any favourable alleles which it has induced. Again, the account of this is rather superficial, and ignores all of the theory which has been developed on the evolutionary modification of mutation rates.

The least radical possibility is that phenotypic variation that is not normally visible can be generated by an environmental extreme, allowing selection to pick out underlying genotypes that are favoured under the extreme conditions. This is, of course, at the core of Waddington’s idea of ‘genetic assimilation’. Hoffmann and Parsons review experimental data on this phenomenon, and reasonably but rather limply conclude that ‘it is difficult to determine how important this process has been in natural populations’ (p. 52).

Later chapters deal with less controversial topics: selection in extreme environments, limits to adaptation, tests of whether predicted patterns of response to environmental changes are observed, and the role of environmental change in causing extinction. The book concludes with discussion of conservation issues.

While this book reviews a multitude of facts, many of which are extremely interesting, I doubt that the critical reader will be convinced that there is really something exceptional about extreme environments, and that the properties of such environments require extensive revisions to conventional notions about evolution. Indeed one of the problems in adopting such a position is that it is unclear what is really meant by an extreme environment. Unless a change in the environment lasts for a sufficiently long time that there can be an evolutionary response to it, it will not cause the evolutionary transformation of a species, although it may cause its extinction. If it does last for a long enough time, then the environment can only be described as extreme in relation to its previous state. It thus becomes quite arbitrary as to where to draw the line between ‘normal’ and ‘extreme’ environmental change. The relative rapidity of evolutionary change induced by the latter may make it easier to tell what is going on in terms of cause and effect, but this does not necessarily imply that there is anything special involved, nor that most cases of profound evolutionary change have involved responses to very sudden environmental changes rather than more gradual ones. The book also suffers from a lack of intellectual coherence; the authors tend to pile fact upon fact, and to sit on the fence on controversial issues, rather than making a clear-cut case for or against a given hypothesis. There is an almost total disregard of theoretical work. For these reasons, I can be only lukewarm about the book.

BRIAN CHARLESWORTH
Institute of Cell, Animal and Population
Biology, University of Edinburgh

The Molecular and Cellular Biology of the Yeast Saccharomyces. Volume 3: Cell Cycle and Cell Biology. Edited by John R. Pringle, James R. Broach and Elizabeth W. Jones. Cold Spring Harbor Laboratory Press 1997. 1131 pages. Price \$115 (cloth) \$75 (paper). ISBN 0 87969 356 8 (cloth) 0 87969 364 9 (paper).

This long-awaited volume completes the update of the 1980s volumes 'The Molecular Biology of the Yeast *Saccharomyces*' which have proved themselves such a useful resource to researcher, teachers and students. Yeast research has moved on dramatically since the first edition, and indeed since the first two volumes of the present series appeared in the early 1990s. It is pertinent to ask whether a monograph that concentrates on a single type of organism remains a worthwhile endeavour, in view of the large body of observations from the past decade showing that, to a first approximation, all eukaryotic cells are basically the same. Yeasts have their advantages as experimental systems – primarily their genetic manipulability – but the question remains: does it make sense to review general cell biological processes on a single organism basis? More specifically, is a book of this sort, comprising a number of such reviews, a useful resource? I return to these questions at the end of the review.

As usual with multi-author volumes, not only the subject material but also the approaches and styles of the chapters are varied, reflecting the views and backgrounds of the authors. Thus the cytoskeleton chapter (Botstein *et al.*) contains an excellent and comprehensive discussion of various genetical methods that have been (or that could be) used to identify new gene functions related to known genes: this chapter should be compulsory reading for research students about to embark on a new genetic screen! One surprising conclusion of this chapter is that in *Saccharomyces*, the only essential role of microtubules appears to be in the mitotic spindle, a striking contrast with other cell types where other roles have been established. A general question raised is whether we now understand the basics of the roles of the main cytoskeletal components to the point where future work will be largely infilling, or whether major new insights are likely.

The chapter by Wenthe, Gasser and Caplan is really a composite. The first part deals with the structure of chromatin in *Saccharomyces* and how this relates to the nucleoskeleton, how chromatin structure is affected during mitosis, and the differences in structure found at the specialized telomere and centromere regions. There are useful comparisons here with the corresponding structures in *Schizosaccharomyces*, which have been less intensively studied, though the gap is closing. This first section is fairly independent

of the latter two parts of this chapter, which together form a coherent unit dealing with the nuclear envelope, and nuclear import and export. The nicely illustrated section describing the behaviour during mitosis of the nuclear envelope, spindle pole bodies and microtubules complements the discussion of similar material in the Botstein *et al.* chapter, and connects up well with the more specialized chapter by Marsh and Rose.

There are three chapters on aspects of the mitotic cell cycle: one on *Saccharomyces* and two on *Schizosaccharomyces*, the cell cycle being one area where fission yeast has at least held its own against the distantly related budding yeast. While each of the three chapters stands well on its own, I feel that a chapter or part of a chapter comparing the rather different cell cycles of the two yeasts, and including a comparison with higher cells, would be welcome.

The *Saccharomyces* chapter, by Lew *et al.*, has an almost philosophical structure based around a central (and often forgotten) question about the nature of the cycle: does it reflect the operation of a master-clock or is it a sequence of dependent events? (The answer is yes, for readers unfamiliar with this field). The description of how the different *Cdc28* activities are integrated within the context of the whole cycle is exceptionally clear and logical, no easy task for a field of this complexity. Finally, I found the deconstruction of the classical view of cell cycle phases and the redefining of the cell cycle highly thought-provoking.

The *Schizosaccharomyces* chapters on cell cycle control (MacNeill and Nurse) and mitosis and cytokinesis (Su and Yanagida) are valuable contributions of rather different types. The MacNeill and Nurse chapter is focussed around the regulatory controls that act in the cycle, with the central role of the Cdc2 kinase a thread that runs through discussions of the regulation of start, the initiation of DNA replication, and the G2-mitosis transition. A final section deals with dependency relationships in the cell cycle – why must DNA replication normally precede mitosis, why is the nuclear DNA replicated only once per cycle, and so on. The thoughtful discussion about the checkpoint mechanisms that underly these dependencies brings together a good deal of published material, and forms an excellent review of this increasingly complex area.

The chapter by Su and Yanagida is primarily concerned with mitosis and cytokinesis in fission yeast. Mitosis and its regulation involves the co-ordinate activity of a large number of proteins and protein complexes, and it is a tribute to the authors that their presentation is both clear and comprehensive. I found their diagrammatic representations of cellular processes and mutant phenotypes very helpful in this context. In addition to mitosis and cytokinesis, the chapter includes useful sections on cell morphogenesis and aspects of chromosome structure,

as well as a little on cell cycle control that overlaps with the MacNeill and Nurse chapter. It is unfortunate that the chapter appears to have been finalized some time before publication and that consequently some exciting recent developments such as those concerning the anaphase promoting complex are not discussed.

Three chapters are devoted to the sexual activities of yeasts: for *Saccharomyces*, conjugation and meiosis/sporulation are dealt with separately, while all aspects of *Schizosaccharomyces*, sexual development are covered in a single chapter. This nicely reflects the biology of the two yeasts: in *Saccharomyces* the conjugation of two haploids to form a diploid cell, and the events of meiosis and sporulation that complete the sexual cycle, are quite distinct events requiring different sets of genes and environmental conditions. In *Schizosaccharomyces*, meiosis normally follows immediately after conjugation, and the genetic and environmental requirements for the two processes overlap extensively. As pointed out in the *Schizosaccharomyces* chapter by Yamamoto *et al.*, there is little similarity in the regulation of sexual differentiation in the two yeasts, nor are there many homologies between the components involved. Research into the mechanisms of cell and nuclear fusion is much further advanced in *Saccharomyces*, and the fascinating chapter by Marsh and Rose shows quite how many aspects of cellular metabolism, including several described in other chapters, are involved in these processes. The regulation and events of meiosis, including recombination and sporulation, are described in a thorough chapter by Kupiec *et al.*, again bringing in many aspects of cell regulation, metabolism and morphogenesis.

To return to the questions posed earlier, there is no question but that this volume contains a large amount of information and useful discussion about the yeasts

Saccharomyces and *Schizosaccharomyces*. The lists of relevant genes and gene products tabulated and referenced in several of the chapters are valuable resources in their own right. Reviews similar to those included here are to be found elsewhere, but for libraries, yeast workers and those just entering the field of yeast research, a comprehensive book like this represents a good investment. Cross-referencing between chapters stresses the inter-relatedness of many aspects of yeast biology, and the editors have done a good job in encouraging this. On the downside, the rate of research progress, not only in yeasts but in other organisms, means that this book will date rapidly, and, as mentioned above, some chapters are already significantly behind the cutting edge. One particular aspect not discussed is how the availability of the entire yeast genomic DNA sequence is affecting the way in which we conduct research, presumably because the sequencing project was completed too late.

Although the title of this book refers only to *Saccharomyces*, three of the 12 chapters deal with *Schizosaccharomyces*. The significant proportion of non-*Saccharomyces* material reflects a change in the nature of cell biology research: while at the time of the first edition, yeast (*Saccharomyces*) was being promoted as the model eukaryote, differences among organisms are now as apparent as similarities. Comparative studies on *S. cerevisiae* and its very distant cousin *S. pombe* have been revealing about what is conserved and therefore likely to be of fundamental mechanistic and evolutionary significance. I therefore look forward to a third edition of this series with major changes in both title and content.

P. A. FANTES

*Institute of Cell and Molecular Biology
University of Edinburgh*