Book review

A Primer of Genome Science

A Primer of Genome Science, 2nd Edition. By Greg Gibson and Spencer Muse. Sinauer Associates, Sunderland, MA, 2004. 378 pp. $59.95

Genome science integrates high throughput DNA sequencing, gene expression and function studies, and bioinformatics to understand the structure, content and evolution of genomes. This different approach has begun to influence fields as diverse as medicine, comparative physiology, ecology and evolutionary biology. In their second edition of *A Primer of Genome Science*, Gibson and Muse (2004) provide a current and highly readable summary of this new and rapidly expanding field, primarily targeted for advanced undergraduate and beginning graduate students interested in genomic research. The text, which is clearly written and complemented by a number of well-chosen illustrations, makes complex concepts more accessible to the novice molecular biologist. Supplementary information, presented in boxes throughout the book, provide excellent short summaries of several issues that can be explored in more detail through the primary literature. The utility of this book is further enhanced by the inclusion of exercises that provide hands-on instruction in genomic techniques and databases.

Although this book is written as an introduction to genome sciences, there are several sections that are relevant to the practicing systematist, particularly if he or she is interested in taking advantage of the large amount of information now being gathered by the expanding community of genomic researchers. The first chapter is an overview of the major genome projects currently underway or recently completed. While this chapter may have been written to focus on only a few model organisms (the usual suspects of human, "mouse", and "fly" come to mind), Gibson and Muse have made an effort to cover the diversity of genome projects. The increasing taxonomic breadth of genome projects, combined with initiatives like Assembling the Tree of Life (AToL), will help stimulate more comparative studies by those working on species related to model organisms. Such an approach will have the symbiotic effect of benefiting genome assembly and annotation through methods like phylogenetic shadowing (Boffelli et al.,

2003) while providing systematists with more characters for phylogenetic inference. This chapter includes an overview of GenBank files (box 1.2) that will be helpful to those who are interested in designing new primers for phylogenetic analyses. There are also exercises (1.3) designed to teach the reader how to compare functional and structural units of homologous gene regions, tools that would be useful to the systematist interested in aligning, partitioning, and weighting data based on functional and/or structural parameters.

Chapter two deals with the methodology of high throughput genome sequencing and the functional annotation of completed genomes. Molecular systematists will be familiar with the sequencing strategies and will find this part a review of modern techniques. The latter half of this chapter, however, deals with building gene and gene family phylogenies to aid in the annotation of functional classes of genes (Eisen, 1998). This is a fascinating application of phylogenetic methods that will no doubt synergistically influence the number of gene regions available to systematists in the future. While this chapter is very well written and serves as a good brief overview, it also includes some sections that systematists are likely to find inadequate. The sections on sequence alignment (pages 74–79) do not go into nearly enough detail for researchers who rely on homology statements to build phylogenetic trees. Furthermore, hidden Markov, Bayesian and direct optimization (Wheeler, 1996, 1999) methods of alignment are not reviewed. This is likely due to the target audience, but adding a section on "Further Reading" seems appropriate for future editions. The phylogenetic analysis section, while short (pages 116–121), provides an excellent overview of parsimony, likelihood, and distance methods (Bayesian methods are not covered) for a molecular biologist. Again, however, this section is too brief to be of use for anyone interested in more detail and additional reference material (Hillis et al., 1996; Schuh, 2000; Felsenstein, 2004) would be of great benefit. A number of other relevant sections, such as the overview of BLAST (box 2.2) will benefit those wishing to mine data from GenBank. However, more detail on interpreting the output of BLAST searches and on some of the search functions available through the

NCBI website (taxonomy report, taxonomy browser) would be useful for systematists.

The third chapter focuses on the nature and detection of single nucleotide polymorphisms (SNPs). Perhaps the most interesting part of this chapter deals with high throughput methods of SNP discovery, detection and genotyping, many of which can be done cheaply and rapidly without sequencing. This section may be more applicable to population geneticists than systematists, but those working close to the species level may find this chapter applicable to their own research. For example, box 3.2 provides an excellent short overview of the coalescent, a body of theory that may have implications for phylogeny reconstruction close to the species level (reviewed in Felsenstein, 2004).

Chapters four and five are geared to more functional analyses of gene expression and proteomics and do not at first glance contain much of interest to the systematics community. However, genome researchers commonly use clustering algorithms to organize gene expression (fig. 4.12) and protein interaction (box 5.2) data. It is possible that these types of analyses can be improved through the incorporation of cladistic techniques (Planet et al., 2001). It is also possible that shared gene expression profiles may provide a useful set of characters when incorporated into phylogenetic analyses.

Chapter six is a prospectus for future research on genomes and their evolution. The authors envision an approach that will include the integration of genome science with ecology and evolutionary biology, as well as with other fields. Phylogenetic systematics will be an invaluable framework for genome scientists, not only to examine historical, structural, and functional aspects of genomics, but also to pose and test hypotheses of comparative genome evolution. Likewise, it is exciting to consider building phylogenetic trees based on whole genomes of a given group of organisms. Although much computational and theoretical work remains to be done in this area, a phylogenetic approach will doubtless be of benefit to both systematists and genome scientists in the future.

References

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., Rubin, E.M., 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science, 299, 1391–1394.

Eisen, J.A., 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res. 8, 163–167.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

Gibson, G., Muse, S., 2004. A Primer of Genome Science. 2nd edn. Sinauer Associates, Sunderland, MA.

Hillis, D.M., Moritz, C., Mable, B.K., 1996. Molecular Systematics, 2nd edn. Sinauer Associates, Sunderland, MA.

Planet, P.J., DeSalle, R., Siddall, M., Bael, T., Sarkar, I.N., Stanley, S.E., 2001. Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression. Genome Res. 11, 1149–1155.

Schuh, R.T., 2000. Biological Systematics, Principles and Applications. Cornell University Press, Ithaca, NY.

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics, 12, 1–9.

Wheeler, W.C., 1999. Fixed character states and the optimization of molecular sequence data. Cladistics, 15, 379–385.

Patrick O'Grady
Department of Biology, University of Vermont,
Burlington, VT 05405, USA
Tel.: + 1 802 656 0460
E-mail address: pogrady@uvm.edu