

Ecology, Evolution, and the Genome: A “Whole-Elephant” Reader's Guide

Author: REMINGTON, DAVID L.

Source: BioScience, 54(10) : 950-954

Published By: American Institute of Biological Sciences

URL: [https://doi.org/10.1641/0006-3568\(2004\)054\[0950:EEATGA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0950:EEATGA]2.0.CO;2)

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Ecology, Evolution, and the Genome: A “Whole-Elephant” Reader’s Guide

DAVID L. REMINGTON

The year 2005 will mark the 20th anniversary of the first meetings held to discuss the revolutionary idea of sequencing the entire human genome, heralding the beginning of genomics as we know it today. To date, the primary motivations for conducting and funding genome research have been highly anthropocentric: to expand our understanding of human biology; to improve the diagnosis and treatment of human diseases; and, to some extent, to enhance our ability to feed a growing human population. However, as the field of genomics approaches its third decade, its potential to blossom into a broader discipline that explores fundamental aspects of evolution, ecology, and environmental biology is gaining recognition.

In the past week alone, I have received two notices dramatizing the growing interest in ecological and evolutionary genomics. One is an announcement for the second Gordon Research Conference on Evolutionary and Ecological Functional Genomics, to be held in Oxford, United Kingdom, in the summer of 2005, which addresses “the functional significance of genomic variation for wild organisms (microbial, plant, and animal) in natural biological communities.” The second is for a two-day “Genes in Ecology, Ecology in Genes” symposium sponsored by the Kansas Ecological Genomics Initiative, which defines ecological genomics as a discipline that “seeks to place the functional significance of genes and genomics into an ecological and evolutionary context.” Moreover, a quick database search identified a number of recent papers discussing the use of genomic methods to identify genes and assay gene expression from environmental samples, for purposes ranging from pathogen identification to comprehensive biodiversity surveys (e.g., Call et al. 2003, Letowski et al. 2003, Venter et al. 2004) and even to identification of extraterrestrial life (Cavicchioli 2002).

As a plant geneticist interested in the molecular basis of phenotypic evolution, I find this expanding vision of genomics a welcome and exciting trend. Many questions about the genetic mechanisms responsible for complex trait variation in wild populations, for example, have been largely intractable with the tools available to date. However, adding genomic methods to the tool kits of ecological, environmental, and evolutionary biologists will require investiga-

tors in these fields to become familiar with genomic technologies and associated new approaches to biological investigation. Given the rapid proliferation of the primary literature in genomic research, the need for references that provide “one-stop shopping” for information on various aspects of genomics is critical. Conversely, experts in genome science who may want to delve into ecological and evolutionary questions will need to gain familiarity with the major concepts and topics in these fields. My aim here is to discuss recent books that might be useful to biologists in each of these two groups, with primary emphasis on the former.

Let me begin with two important qualifiers. First, this is not a review of books on evolutionary, ecological, or environmental genomics, for the simple reason that (to my knowledge) no such books have been written yet. Moreover, the accumulated body of knowledge and experimental approaches necessary for such books to be written, let alone useful, are probably several years away. Nevertheless, a number of recent books with genomic and evolutionary biology themes will be useful in preparing readers from various scientific backgrounds to weave the various threads of ecology, evolutionary biology, and genome science into a useful body of research.

Second, I need to qualify what I mean by *genomics* (and the interchangeable term *genome science*). Different scientists use the term *genomics* in such different ways that one is reminded of the proverbial blind men describing an elephant. For one researcher, genomics seems to be synonymous with whole-genome sequencing and analysis, while another equates it with microarray analysis of gene expression. Yet another investigator may be referring to expressed sequence tag-based “gene discovery” projects. My best attempt at describing the whole elephant is to characterize genomics as the study of biological phenomena at the scale of the entire genome, with a level of resolution corresponding to individual genetic loci.

David L. Remington (e-mail: dlreming@uncg.edu) is an assistant professor in the Department of Biology, University of North Carolina, Greensboro, NC 24702. He conducts research on the genetic mechanisms of phenotypic evolution in several plant lineages, including *Plantago lanceolata* and the Hawaiian silversword alliance. © 2004 American Institute of Biological Sciences.

This definition is intentionally broad enough to encompass elementary tools such as genomewide genetic mapping studies and, on the other hand, “postgenomic” approaches such as proteomics. At the same time, it necessarily excludes both traditional gene-by-gene molecular analyses and summary-statistic approaches to characterizing whole-genome effects, such as traditional quantitative genetic analysis.

Many readers will undoubtedly be interested in a comprehensive, easy-to-read introduction to genomics that addresses both the experimental approaches and the technologies that are used. *A Primer of Genome Science* (2002), by Greg Gibson and Spencer Muse, is still probably the most useful and accessible book in this category. Gibson and Muse, both professors at North Carolina State University, primarily set out to write a textbook for courses in genome science, and I use it as such in the genomics course I teach to upper-division undergraduates and graduate students. However, the book’s style and content make it equally well suited for more established scientists who want an informative overview of genomics.

Much of the book’s utility stems from the fact that Gibson and Muse themselves come from outside the mainstream of the genomics community: Gibson specializes in developmental quantitative genetics, while Muse studies analytical methods in molecular evolution. Perhaps from this perspective they see the “whole elephant” a little more clearly. After a brief introduction to genome mapping and major completed and ongoing (as of mid-2002) genome projects, the book offers concise discussions of genome sequencing and analysis, gene expression profiling, proteomics, saturation mutagenesis, analysis of single-nucleotide polymorphisms, and metabolic analysis. Attractive graphics illustrate the major themes, and supplemental boxes in each chapter provide clear explanations of the analytical and bioinformatic tools and methods that are essential to the dissemination and analysis of genomic data. It is my fervent hope that the authors will produce regular revisions to keep the book up-to-date on new developments, especially the evolutionary insights being provided by the rapid proliferation of sequenced genomes and the expanding applications of gene expression profiling.

Another useful overview, written from a plant genomics perspective, is *Plant Genomics and Proteomics* (2004), by Christopher Cullis. Cullis, a professor at Case Western Reserve University and a former program director in the National Science Foundation’s Plant Genome Research Program, emphasizes distinctive features of plant genomes and their relationship to unique aspects of plant biology. For instance, a number of genes, encoded in both the nuclear and the chloroplast genomes, encode proteins that regulate photosynthesis and responses to the quality and quantity of light. The immobility of plants, and consequently the extremes of the physical and biotic environments they must

encounter, result in the need for large numbers of genes involved in production of secondary metabolites and pathogen responses. Other idiosyncrasies of plant genomes are extreme variation in nuclear DNA content and the prevalence of genome duplication in polyploidy, both of which bear still-to-be-explored relationships to variability in life history and environment. This plant-specific perspective is welcome, especially since the genome science literature (written mostly by experts in mammalian and microbial systems) is peppered with naive and often patently incorrect characterizations of plant biology.

Cullis begins with an introduction to what has been discovered so far about the structure and organization of plant genomes. His discussion of plant genome evolution clearly illustrates the insights that have been gained from the *Arabidopsis* and rice genome sequences and from more modest efforts in other plant genomes. Evolutionary biologists will undoubtedly take issue with some of Cullis’s statements, such as his definition of orthology and paralogy in terms of gene function rather than evolutionary origin, but I found these to be minor distractions. Subsequent chapters alternate

between discussion of genomic analysis techniques (including sequencing, gene discovery, analysis of gene function, complex trait dissection, and bioinformatics) and description of the biological processes these techniques can be used to investigate. One chapter is devoted to exploring how genomic profiles of gene expression can be used to study the biology of plant-environmental interactions, such as disease and stress responses—a relevant topic for crop breeders that is also of interest to evolutionary ecologists. A chapter on complex trait dissection discusses quantitative trait locus (QTL) mapping, also mainly from a crop science rather than an evolutionary biology perspective. The book ends with a chapter discussing bioethical concerns, in which Cullis addresses the interplay of safety, public perception, trade policy, and regulatory framework surrounding genetically modified crops. On this controversial topic, Cullis appears to have gone to great pains to be informative rather than partisan, although his exasperation at what many crop scientists perceive as uninformed and exaggerated concerns about consumer and environmental safety comes through in places.

Cullis’s writing style is clear and easy to read for the most part, though he does occasionally lapse into jargon, especially when describing bioinformatic databases. He provides reasonably detailed and illustrated explanations of a number of molecular techniques used in genome science, though I find the corresponding explanations in the Gibson and Muse textbook easier to follow.

Genomics becomes an essentially evolutionary undertaking as soon as one starts comparing the content and organization of sequenced genomes from different organisms. Such efforts have been rudimentary in eukaryotes until recently

Genomics becomes an essentially evolutionary undertaking as soon as one starts comparing the content and organization of sequenced genomes from different organisms.

because of the paucity of completely sequenced genomes, especially among closely related pairs of taxa, but are at a much more advanced stage in prokaryotes, of which dozens now have completely sequenced genomes. Two recent books are devoted exclusively to the field of comparative genomics, offering complementary perspectives on the analysis of genome evolution. Cecelia Saccone and Graziano Pesole, of the University of Bari and the University of Milan, respectively, have written *Handbook of Comparative Genomics: Principles and Methodology* (2003) as a comprehensive and detailed reference on comparative genomics. Saccone and Pesole divide their work into three major sections, describing first the major features of prokaryotic, eukaryotic, and organellar genomes; then molecular and analytic methodologies; and finally the major results of comparative genome analyses. The descriptions of genomic databases and computational methods are detailed and moderately mathematical. For example, detailed explanations are given for local and global sequence alignment methods, and for the search algorithms used in BLAST and related programs. The computational methodologies are well illustrated with examples from actual analyses of sequence data. The book is organized to function as a reference volume, and individual chapters stand quite well on their own.

By contrast, *Sequence–Evolution–Function: Computational Approaches in Comparative Genomics* (2003), by Eugene Koonin and Michael Galperin, seems to be designed to be read from cover to cover. If Gibson and Muse approach genomics with an outsider's perspective, Koonin and Galperin, both investigators for the National Center for Biotechnology Information at the National Institutes of Health, are consummate insiders. For Koonin and Galperin, genomics is largely synonymous with sequencing complete genomes and identifying their protein-coding sequences (the process known as annotation), and its primary application is to infer protein function and structure by comparative methods. The result of this perspective is an engaging book that reads like a set of genomic detective stories. Unlike Saccone and Pesole, Koonin and Galperin provide mostly descriptive and nonmathematical explanations of bioinformatic tools and analytical methods. These explanations are filled with detailed examples that illustrate many clever successes, as well as more than a few spectacular failures, in genomic deductions of protein function. Some of their examples are nonbiological, as in one delightful section where they illustrate the principles of sequence alignment using stanzas from Edgar Allan Poe's "The Raven."

Koonin and Galperin's examples of genomic inference of protein function are weighted heavily toward studies in prokaryotes, in accordance with the primary focus of their own research. Not surprisingly, horizontal gene transfer receives center stage in their discussions of evolutionary processes that

have shaped genomes, whereas the gene and chromosomal duplication processes that are more important in eukaryotes get relatively little attention. One concept the authors discuss extensively is that of phyletic patterns, which are the patterns of presence or absence of genes that encode various proteins in different sequenced genomes. Analyses of phyletic patterns reveal instances in which metabolic functions are performed by unrelated genes in different phylogenetic lineages, yielding insights into the evolution of metabolic pathways. Phyletic patterns may reveal enzymes that are present, for example, in pathogenic microbes but not in humans, providing useful targets for new antibiotic drugs. A number of environmental biology applications of this approach can be envisioned for eukaryotes as well (development of more environmentally friendly pesticides comes to mind), but only

when many more complete genomes have been sequenced.

Given its length of 380 pages, a full reading of *Sequence–Evolution–Function* is not a trivial undertaking, and many readers may want to skip some of the more detailed descriptions of bioinformatic software tools or

examples of genomically dissected metabolic pathways. The reward of reading this book, however, will be a greater appreciation for the insights into gene function and evolution that can be gained using comparative genomics. This detective-oriented perspective may be especially useful to those interested in using patterns of gene expression to identify gene regulatory and developmental pathways involved in phenotypic responses. One minor complaint is that poor editing for items such as punctuation detracts from what is otherwise a well-written and engaging book.

Both *Handbook of Comparative Genomics* and *Sequence–Evolution–Function* discuss in some detail the implications of comparative genomics for phylogenetic inference. For example, reconstructing a single "tree of life" may be problematic now that comparative genomics has shown horizontal gene transfer to be such a prevalent feature of evolution, at least in prokaryotes. Koonin and Galperin suggest that genomics will lead to entirely new approaches to phylogenetic analysis. Two ideas they discuss are (1) excluding sequences potentially involved in horizontal gene transfer when estimating divergence for concatenated sets of nuclear DNA sequences and (2) using comparisons of gene order and phyletic patterns as tools for constructing phylogenetic trees. It is clear from the tentative nature of the discussion in these two books that the use of genomic approaches in phylogenetic analysis is still very much in an exploratory phase. Even if these books only stimulate more thinking along these lines, they will have accomplished something.

Microarrays, or "DNA chips," have become the workhorse of genomic studies in recent years. Microarrays are a multipurpose tool that can be designed to compare gene expression in different tissues, environmental conditions, or

Reconstructing a single "tree of life" may be problematic now that comparative genomics has shown horizontal gene transfer to be such a prevalent feature of evolution, at least in prokaryotes.

genotypes (i.e., expression profiling), identify allelic variants in populations, or evaluate the occurrence of genes from different taxa in environmental samples. Embarking on a microarray study, however, is a daunting task that requires careful planning, diligent execution, and expensive and delicate equipment for printing and reading arrays. A recent edited volume, *A Beginner's Guide to Microarrays* (Blalock 2003), provides an introduction to all facets of microarray applications, from array construction to data analysis. The book consists of eight chapters, all written by different authors with expertise in different aspects of microarray technology and analysis. The first four chapters are devoted primarily to the technological aspects of array construction, including surface preparation of the glass slides on which the arrays are printed, preparation of the probes, and printing of the arrays. These chapters have a highly practical focus, with considerable emphasis on the strengths and weaknesses of various alternative technologies for probe and slide preparation and array printing, and on troubleshooting the many items that can (and often do) go wrong in the process. There is quite a bit of redundancy in the material covered in these chapters, but I found the repetition helpful in understanding unfamiliar aspects of microarray technology. These chapters should be useful to researchers interested in designing custom arrays, which are essential for evolutionary biologists working in nonmodel systems that lack "off-the-shelf" arrays from commercial suppliers. A chapter by Robert Searles, of Oregon Health and Science University, discusses in a folksy style the process and pitfalls of setting up a microarray core facility, with insights based on his own sometimes frustrating experiences. Searles's advice is likely to be valuable for others thinking of establishing a similar facility.

For the ecologically minded, a special mention should be made of the second chapter, written by Levente Bodrossy, who addresses the design and use of microarrays to identify microorganisms present in environmental samples. Bodrossy provides considerable detail on software and Web-based tools for aligning gene sequences and designing probes that give desired levels of taxon specificity. This chapter provides a welcome departure from the predominantly biomedical interests of the remaining authors.

In the final four chapters of *A Beginner's Guide to Microarrays*, the emphasis shifts to experimental design, statistical analysis, and data interpretation. Although these chapters cover critical topics for anyone contemplating microarray studies, much of the material will not be accessible to any but the most bioinformatics-literate "beginners." For example, the final chapter, by Willy Valdivia Granda, discusses techniques for clustering gene expression profiles, which are used to infer the functional significance of changes in gene expression, but the amount of undefined technical jargon makes the descriptions extremely hard to follow. Sloppy editing and typesetting are also a major distraction in reading this book. Everyday word processors are generally capable of performing full justification of lines of print without introducing

spaces within words, so it is baffling to see such errors on nearly every page of a commercially published book.

For the reader primarily interested in microarray experimental design and analysis, the second edition of *Guide to Analysis of DNA Microarray Data* (2004), by Steen Knudsen, may be a more useful choice. Knudsen, of the Center for Biological Sequence Analysis at the Technical University of Denmark, provides a detailed but highly readable presentation of microarray image analysis, statistical analysis, clustering, advanced approaches to regulatory network evaluation, and experimental design. Knudsen's mathematical explanations of analytical methods are presented in ways that should be easily understood by the typical nonstatistician.

The books I have discussed so far include little discussion of the relationship between the genome and phenotypic variation. The few exceptions include discussions of associations between genetic polymorphisms and disease susceptibility by Gibson and Muse, and the practically oriented chapter on complex trait dissection in plants in Cullis's book. Since both adaptive evolution and ecological processes depend on phenotypic expression of genome variation, the lack of attention to this topic demonstrates how far evolutionary and ecological genomics lags behind other applications of genome science. However, contributions to another edited volume, *The Evolution of Population Biology* (Singh and Uyenoyama 2004), highlight some of the emerging directions in evolutionary and ecological genomics research. The volume is written in honor of Richard Lewontin, a stalwart advocate for understanding the genome in its developmental and environmental context. While the entire volume is invaluable for striving to synthesize a truly integrative approach to biological inquiry, three chapters are noteworthy for their discussion of genomic themes.

Trudy Mackay's contribution discusses the importance of QTL mapping in bridging the gap between what Lewontin (1974) identified as evolutionarily "uninteresting" Mendelian trait variation and the "unmeasurable" continuous trait variation that is the raw material of evolution. As a genomic tool, genetic mapping has been primarily an initial step toward the real goal of complete genome sequencing. In an ecological and evolutionary context, however, mapping of QTLs to dissect the genetic architecture of trait variation is an important task in itself. QTL analysis can elucidate long-standing questions about the numbers of genes involved in potentially adaptive phenotypic traits, the size of their effects, gene interactions, and pleiotropic effects on other traits and fitness itself. Moreover, as Mackay discusses, QTL analysis is increasingly being used as the basis for finding the individual genes and ultimately the actual polymorphisms responsible for natural variation. QTL analysis has been around for some time, but its use for addressing the nature of adaptation, a challenge raised by Orr and Coyne (1992) more than a decade ago, has been slow to catch on. Mackay's chapter points out the ways in which this is changing, a trend that is due in no small part to her own research program.

In their contribution to *The Evolution of Population Biology*, Daniel Hartl and his coauthors discuss applications of microarray-based gene expression profiling for evolutionary genetics. Comparing genomewide patterns of gene expression across different genotypes, populations, or related species has not been a major use of microarray technology so far, but this approach provides a new and powerful tool for investigating the genetic mechanisms behind trait variation and evolution. Hartl and colleagues provide several illustrative examples of such studies from yeast, *Drosophila*, and nematodes. This is an area of investigation in which great strides are likely over the next few years, as evidenced by other noteworthy studies since the chapter was written (e.g., Rifkin et al. 2003, Schadt et al. 2003, Fay et al. 2004).

A brief chapter by Brian Golding discusses several applications of genome-related bioinformatics tools in population biology. Trends Golding identifies as noteworthy include the use of molecular markers to identify species and even individual populations in biodiversity surveys, the use of comparative genomics to study the spread of antibiotic resistance and environmental adaptations in bacteria, and the use of microarrays to identify gene expression patterns associated with exposure to environmental hazards.

This special book article would be incomplete without some discussion of a book that has little to do with genomics but much to do with investigations of evolutionary biology. The newly released second edition of John Avise's *Molecular Markers, Natural History, and Evolution* (2004) provides a detailed and comprehensive presentation of the scope of molecular analysis of evolutionary topics, ranging from population-level dynamics to attempts to reconstruct the tree of life. In revising his original 1994 edition, Avise has chosen to go deeper rather than broader with respect to molecular tools, emphasizing new evolutionary insights from existing molecular marker techniques rather than newer technologies. Consequently, other than some discussion of comparative genomic insights into macroevolution, there is relatively little reference to genomic approaches in this book. Even QTL analysis is barely addressed, and Avise emphasizes its limitations as much as its potential value. By contrast, Avise's coverage of the kinds of evolutionary questions that are being addressed with molecular tools is expansive. In the first part of the book, he discusses long-standing questions about the nature of the forces shaping genetic variation and DNA sequence evolution. This discussion is followed by a summary of protein- and DNA-based molecular markers and analysis methods. He then devotes entire chapters to molecular analyses of topics such as population structure and phylogeography, speciation and hybridization, phylogeny, and conservation genetics.

Avise's book is of value for experts in evolutionary biology and genomics alike, for the simple reason that it provides such thorough coverage of the kinds of evolutionary questions that could prospectively be addressed by genomic approaches. If ecological and evolutionary genomics is to emerge as a full-fledged field of investigation, then genomic approaches will need to be brought to bear on the broad range of ecological and evolutionary questions that are currently being studied with more basic molecular tools. Perhaps then we will be in a better position to consider ways in which the full range of genomic tools—including genomic and expression libraries, genome sequencing, and gene expression profiling—can be used to gain new insights into these questions. I believe the genomic frontiers in evolutionary biology and ecology will prove to be at least as fascinating as those in the biomedical sciences as genomics enters its third decade.

References cited

- Avise JC. 2004. *Molecular Markers, Natural History, and Evolution*. 2nd ed. Sunderland (MA): Sinauer.
- Blalock EM, ed. 2003. *A Beginner's Guide to Microarrays*. Boston: Kluwer Academic.
- Call DR, Borucki MK, Loge FJ. 2003. Detection of bacterial pathogens in environmental samples using DNA microarrays. *Journal of Microbiological Methods* 53: 235–243.
- Cavicchioli R. 2002. Extremophiles and the search for extraterrestrial life. *Astrobiology* 2: 281–292.
- Cullis CA. 2004. *Plant Genomics and Proteomics*. Hoboken (NJ): Wiley-Liss.
- Fay JC, McCullough HL, Sniegowski PD, Eisen MB. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biology* 5: R26.
- Gibson G, Muse SV. 2002. *A Primer of Genome Science*. Sunderland (MA): Sinauer.
- Knudsen S. 2004. *Guide to Analysis of DNA Microarray Data*. 2nd ed. Hoboken (NJ): Wiley-Liss.
- Koonin EV, Galperin MY. 2003. *Sequence–Evolution–Function: Computational Approaches in Comparative Genomics*. Boston: Kluwer Academic.
- Letowski J, Brousseau R, Masson L. 2003. DNA microarray applications in environmental microbiology. *Analytical Letters* 36: 3165–3184.
- Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.
- Orr HA, Coyne JA. 1992. The genetics of adaptation: A reassessment. *American Naturalist* 140: 725–742.
- Rifkin SA, Kim J, White KP. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genetics* 33: 138–144.
- Saccone C, Pesole G. 2003. *Handbook of Comparative Genomics: Principles and Methodology*. Hoboken (NJ): Wiley-Liss.
- Schadt EE, et al. 2003. Genetics of gene expression surveyed in maize, mouse, and man. *Nature* 422: 297–302.
- Singh RS, Uyenoyma MY, eds. 2004. *The Evolution of Population Biology*. Cambridge (United Kingdom): Cambridge University Press.
- Venter JC, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74.



Your antidote to information overload

Conservation In Practice is an all-in-one resource that will keep you abreast of the trends, the people, and the research that are pushing the boundaries of the field.

It's short and to the point. In every issue, you'll find reports that are factual, bold, controversial, and always personal and accessible.



www.conbio.org/inpractice

**Subscribe
Online!**