

Book Selection

Molecular Systematics

Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717, USA

Review of David M. Hillis and Craig Moritz, eds. *Molecular Systematics*, 1990, Sinauer Associates, Sunderland, MA. xvi + 588 pp.

“His [Matisse’s] art spoke of joy and happiness almost beyond human reach; it carried the dream of an ever-present golden age and a promise of beauty for tomorrow.”—Françoise Gilot, *Matisse and Picasso. A Friendship in Art*, Doubleday, New York, 1990, p. 5.

The notion that living organisms can be divided into two kingdoms, animals and plants, prevailed from Greek antiquity until recent times. The existence of five kingdoms became accepted in the 1960s, following Whittaker’s (1959) proposal that three additional kingdoms should be recognized: fungi, unicellular eukaryotes, and prokaryotes. Woese and colleagues (1990) now argue that the five-kingdom system is unsatisfactory because it assigns identical rank to five groups of organisms that are unevenly different from one another. They propose a new taxonomic category, “domain,” of rank above kingdom, and that there be three domains: Eucarya (including all eukaryotes), Bacteria (including the eubacteria), and Archaea (including the archaeobacteria, a name they recommend be abandoned as it suggests a false phylogenetic relationship between the Archaea and the Bacteria).

Woese et al.’s proposal is buttressed by the extent of nucleotide divergence in the genes coding for rRNA, but much more than this divergence is implied. The rRNA genes evolve very slowly and hence provide reliable clues of remote evolutionary events. The rRNA genes indicate that the Bacteria and the Archaea diverged in the remote recesses of the evolutionary past and that the eukaryotes came about as an offshoot of the same evolutionary branch that gave rise to the Archaea. The Bacteria are no less

divergent from the Archaea than either of these from the Eucarya; hence, their segregation into two taxa of equal rank. The proposal for three domains with equal taxonomic rank reflects an effort to bring taxonomic classification into line with phylogenetic information.

Mayr (1990), the doyen of American systematists, has retorted, apparently much to the surprise of Woese et al. (1991), that these authors’ scheme replaces an imbalance by a different one, because it ignores the enormous evolutionary step from the prokaryotes to the eukaryotes. Mayr proposes instead two domains, each with two subdomains: Prokaryota (subdomains Eubacteria and Archaeobacteria) and Eukaryota (subdomains Protista and Metabionta). Woese et al. (1991) have responded if not quite by crying foul at least by stating that Mayr’s proposed classification is “artificial” and returns to “the flawed conventional view that divides the living world into eukaryotes and prokaryotes.”

It is not my intention here to enter this dispute. I rather cite it because it illustrates two important themes of *Molecular Systematics*, a book that is as thoughtful as it is useful. One theme is the amplitude of new knowledge that has already come about from the application of molecular biology to the study of evolution, with much, much more to come. The other theme is that empirical observations by themselves do not advance knowledge or settle issues, although they can be used for testing hypotheses and for inspiring their formulation. Scientific knowledge advances by the corroboration or rejection of hypotheses. But hypotheses are theoretical constructs. Hypotheses guide observation and experiment because they suggest what to observe. It is the imaginative conjecture of what might be true that pro-

vides the incentive to seek the truth and a clue as to where we might find it (Medawar 1967). In the case of taxonomy, hypotheses proposing certain relationships can only be tested within the context set by a higher-level theory that defines a particular system of classification.

The disagreement between Mayr and Woese et al. emerges from two different theories of taxonomy. For Woese and colleagues, as it is for cladists, phylogenetic history is all-prevailing in determining taxonomic classification. For Mayr, other biological considerations, such as morphological complexity, should be taken into account as well. A cladistic taxonomist might include crocodiles and birds in the same taxon, different from a sister taxon that would include other reptiles such as the turtles. This, I presume, would not be acceptable to Mayr for analogous reasons as those he levels against Woese et al.'s proposal.

The sequence of nucleotides in the DNA contains genetic information very much in the same way as the sequence of letters of the English alphabet in a book contains semantic information. The genetic information in the DNA directs the development and function of the organism, but it is also a record of the evolutionary history of the species. Molecular biology has made it possible to access that record directly, whereas only the entangled outcomes of development and function were accessible before. Direct access to genetic and evolutionary information provides a new and very informative level for making comparisons between organisms, the mainstay of phylogenetic studies.

The powerful methods of Mendelian analysis had made it possible to compare individuals genetically, but only those of the same species and only in organisms that could be experimentally bred in the laboratory or in the field. With the advent of molecular biology, these two restrictions disappeared. It is not only that genetic causation can be ascertained without the laborious process of breeding progeny, but also that genetic homologies can be recognized between species, even remotely related, that cannot interbreed. This possibility brought into fruition the discipline of evolutionary genetics that had previously existed in rudiment. At the same juncture, new vistas appeared in the horizon of other evolutionary disciplines, systematics being one of them. This has elicited a euphoria in evolutionary biology because so many new problems can be addressed and old ones settled: "the dream of an ever-present golden age."

Systematics is the study of biological diversity. Its subject may conveniently be differentiated into two: population structure (below the species level), which includes such matters as mating system, genotypic variation, and geographic subdivision; and

(above the species level) phylogeny, the study of evolutionary history, with its bearing on taxonomic theory and practice. *Molecular Systematics* dedicates six chapters, one to each of the molecular techniques most extensively used in systematics: isozymes, immunology, cytogenetics, DNA-DNA hybridization, restriction site analysis, and DNA (and RNA) sequencing. (Protein sequencing was of great historical importance in molecular systematics, but it is omitted on the grounds that it has been largely replaced by nucleic acid sequencing.) Some techniques, such as isozyme and restriction site analysis, are useful for investigating population structure as well as the phylogeny of organisms evolutionarily not very distant. DNA-DNA hybridization and immunology also are helpful for studying the phylogeny of organisms not very distant, but are not suitable for investigating population structure. DNA and RNA sequencing provide the most detailed information for all sorts of problems, although sequencing may be prohibitively expensive for investigating issues of population structure. A table on p. 504 of *Molecular Systematics* provides a bird's-eye view of the range of applications for each technique.

The six chapters on molecular techniques are how-to guides and much more. Each chapter starts with a clear formulation of the principles involved, the various methodologies and their assumptions, their applications and limits; laboratory setups, protocols, interpretations, and recipes are described with detail sufficient for a person who has never used a particular methodology to get going, but also for improving the practice of those who may have already used that methodology. There is nothing critical I want to say about any of these chapters as they are on the whole extremely successful. Much credit is due to the various authors, but also to the book editors, who obviously provided well-defined outlines and criteria and have exercised strict editorial control for quality, consistency, and completeness.

Not as a criticism but rather as testimony of how fast the discipline is advancing, I will call attention to the omission of the recently developed method of random amplified polymorphic DNA (RAPD) (Williams et al. 1990; Welsh et al. 1991). This method, based on polymerase chain reaction amplification with short DNA (10 bp or a few more) oligomer primers of arbitrary nucleotide sequence, can yield genetic maps of substantial resolution (say 100 markers) for a previously completely unmapped organism in just a few months of work. The applications of the RAPD method will surely multiply in the forthcoming months and years, not only in genetics, but also in systematics, for investigating issues such as linkage disequilibrium, population variation, and phylogeny.

Molecular Systematics includes two chapters (first and last) that provide conceptual context and comparative guidance, two chapters on sampling design and sample collection and preservation, and two chapters on data analysis. These last two all-important chapters properly emphasize that data analysis and interpretation must be guided by biological models, which define the hypotheses that are being tested. Both chapters are exemplary. Weir's chapter on intraspecific variation is a model of clarity, succinctness, and practicality.

The chapter on phylogeny reconstruction by Swofford and Olsen is, at 90 pages, the longest in the book. The authors provide a cool-headed and unopinionated analysis of a field that is plagued with controversy and acrimony. They reiterate the need to distinguish, for any phylogenetic method, between the optimality criterion (the theoretical framework and conceptual assumptions) and the algorithm (the recipe of computational steps for obtaining the phylogeny). The properties of discrete characters and distance data are delineated; and the various methods are described that rely on one or the other type of data. As they announce, the authors "focus on methods that are currently in widespread use or that are likely to be used in the foreseeable future" (p. 411) and they do so with as much analytical and practical detail as would seem feasible within the space constraints. One method not included is the statistical geometry in distance, or sequence, space of Eigen and colleagues (Winkler-Oswatitsch et al. 1986; Eigen et al. 1988, 1989; see Maynard-Smith 1989). This method is not "currently in widespread use," and it remains to be seen whether it will be much used in the foreseeable future, so that Swofford and Olsen are not omitting anything they would be expected to deliver. The method of statistical geometry provides simple rules for selecting the appropriate phylogeny and for deciding whether a tree-like topology is more likely than a "bundle" (i.e., an effectively simultaneous split of multiple branches) or a "net" (such as might arise by hybridization or lateral genetic transfer). This method has been applied to the evolution of tRNA and rRNA molecules (Winkler-Oswatitsch 1986; Eigen et al. 1989), but to my knowledge it has not yet been used to determine the phylogeny of a set of organisms. The statistical properties of the method are far from well defined, but it has con-

ceptual and (relatively speaking) computational simplicity; and it may be particularly helpful for discriminating between parallel or superimposed changes on the one hand and lateral genetic transfer on the other. The excellent chapter by Swofford and Olsen might have been a good vehicle to introduce statistical geometry to many molecular evolutionists and provide guidance.

There is little negative that would be appropriate to say about *Molecular Systematics*. There are infelicities of style here and there and occasional lapses into obscurity. One wonders what muse inspired the authors to write without any trace of irony (and distracted the editors from red-penciling) the following: "The best way to avoid random errors is to obtain an *infinite* amount of data. . . . Since this option is *unavailable to most researchers* . . ." (p. 491, my italics). My intention, however, is rhetorically to switch this reproof into the cue for a commendatory remark: the searching eye of the reviewer fails to turn almost anything that might seem faulty in this extraordinary book.

References

- Eigen M, Winkler-Oswatitsch R, Dress A (1988) Statistical geometry in sequence space: a method of quantitative sequence analysis. *Proc Natl Acad Sci USA* 85:5913-5917
- Eigen M, Lindemann BF, Tietze M, Winkler-Oswatitsch R, Dress A, Von Haeseler A (1989) How old is the genetic code? Statistical geometry of tRNA provides an answer. *Science* 244:673-679
- Maynard-Smith J (1989) Trees, bundles or nets? *Trends Ecol & Evol* 4:302-304
- Mayr E (1990) A natural system of organisms. *Nature* 348:491
- Medawar PB (1967) *The art of the soluble*. Methuen, London
- Welsh J, Petersen C, McClelland M (1991) Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. *Nucleic Acids Res* 19:303-306
- Whittaker RH (1959) On the broad classification of organisms. *Q Rev Biol* 34:210-226
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531-6535
- Winkler-Oswatitsch R, Dress A, Eigen M (1986) Comparative sequence analysis. *Chem Scr* 26B:59-66
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576-4579
- Woese CR, Kandler O, Wheelis ML (1991) A natural classification. *Nature* 351:528-529