CURRENT SCIENCE

Volume 79 Number 11 10 December 2000

EDITORIAL

Systematic biology in the information age

The fusion of biology and information technology has spawned a new field, bioinformatics. All of a sudden, governments, venture capitalists and science adventurers see a fortune to be made, in a discipline whose origins are so recent, that even its definition is somewhat blurred. Bioinformatics grew in the shadow of the genomics revolution in the 1980s and 1990s, with the dramatic developments in the areas of information storage, retrieval and transfer providing spectacular support. The flowering of the Internet and the successful completion of major genome sequencing projects have happened at almost the same time, propelling bioinformatics to the forefront of scientific disciplines. There is presently, an enormous volume of nucleic acid sequence information pouring into computerized databases, with the raw information requiring considerable refinement and analysis before useful biological interpretations can be formulated. The term 'data mining', which is frequently used at meetings, refers to the fond hope that bioinformaticists, like miners of yore, will pull out nuggets of virtual gold from the enormous rubble of accumulated DNA sequence information. There are other areas of bioinformatics too. Crystallographers, and more recently NMR spectroscopists, have been depositing three-dimensional structural information on biological molecules, in the form of positional coordinates for individual atoms, into public domain data banks. While less extensive than sequence data, even the 3D-databases are growing quickly, facilitating statistical analysis and pattern recognition forays. The newly developing field of structural genomics soon promises to fill computer banks with structures of protein molecules of unknown function; truly turning biochemistry on its head. The remarkable speed with which DNA expression profiling has progressed, signals the arrival of another flood of biological information, this time on differential gene expression in diverse settings. In all these cases, the power of modern computers and the efficiency of global data transfer are central to further developments.

While molecular research has provided the rationale for the marriage of information technology and biology, there are clearly other areas of biological research, which will soon be swept into the digital age. A recent issue of *Science* (29 September 2000), highlights the potential importance of information technology for biodiversity research. Issuing a call for a 'global biodiversity map', Edward O. Wilson, the high priest of sociobiology, who coined the term 'biodiversity', argues that 'to describe and classify all of the surviving species of the world deserves to be one of the great scientific goals of the new century'. Wilson and others guess that the number of species of organisms that inhabit the earth may be of the order of 10 million; with about 1.5–1.8 million having been scientifically identified and named, in two centuries of 'the Linnaean enterprise' (Wilson, E. O., *Science*, 2000, **289**, 2279). Central to this goal of complete classification would be an army of expert taxonomists and systematic biologists, who have unfortunately been marginalized in biology departments across the world as molecular approaches grew from strength to strength. Wilson emphasizes this, when he argues that the 'key choke point . . . is not the tools of informatics, most of which are already at hand. Rather, it is the severely limited capacity of museums and other collections-oriented facilities to collect, prepare and analyse specimens, and the shortage of expert taxonomists to do the job'. In recognizing that taxonomists may indeed be an endangered species, the U.S. National Science Foundation has funded an initiative a few years ago, called the 'Partnerships to Enhancing Expertise in Taxonomy', whose purpose will eventually be to create a new biologist, 'comfortable with molecular and computer tools as well as with the microscope and collecting kit' (Pennisi, E., *Science*, 2000, **289**, 2306).

In India both biodiversity and bioinformatics are buzzwords; cash cows to be milked in the meeting rooms of funding agencies. Bioinformatics, is largely confined to setting up centralized sites for data dissemination, with little by way of original research ('mining'), at centres across the country. Indeed, most bioinformatics units seem content to be 'mine owners', still waiting for the 'miners' to descend into the pits to hunt for gold. Since very little by way of original data is received from local sources, these centres are repositories for international public domain databases. Biodiversity research is also hampered by the gradual disappearance of expert taxonomists. Conservation biology focuses on the few visible and publicly appreciated species, which provide a rallying point for activists; the millions of organisms that remain unseen and unidentified are of little concern. While our 'biodiversity resources' are trumpeted at every forum, the magnitude of the task of even finding out what we have is not easily appreciated. Bioinformatics provides powerful tools. Biodiversity research will allow us to realize the richness of our biological resources. But, at present, the prospect of moving purposefully towards the Wilsonian goal of a complete classification in India appears remote. Such projects require a level of organization, commitment and staying power that may be hard to come by in our midst. But, as a first measure we may indeed begin by seeding a 'taxonomic revival', restoring systematic biology to its proper place. Ensuring the integrated teaching of biology in a fashion that draws the connections between systematics and molecular approaches, emphasizing the tools of the information age, may be a good starting point. Finally, in considering the role of 'informatics' in biology, it may be prudent to remember the fine distinctions between information, knowledge and wisdom.

P. Balaram

Ayan Guha