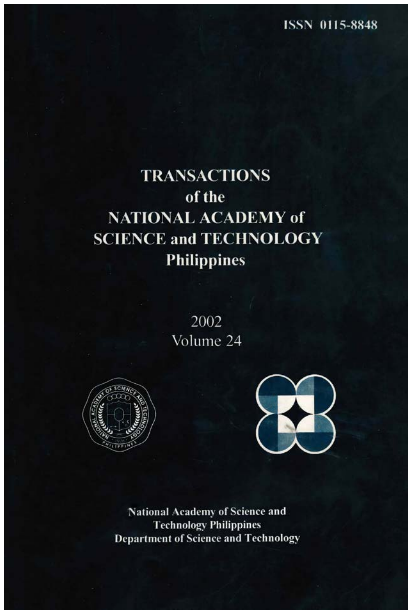
TRANSACTIONSNASTPHL

ISSN 0115-8848 (print)
ISSN 2815-2042 (online)

https://transactions.nast.ph
Vol. 24 Issue No. 2 (2002)
doi.org/10.57043/transnastphl.2002.5086



Bioinformatics: Genomics and Beyond (A User's Perspective)

Nina Rosario L. Rojas

Department of Chemistry, School of Science & Engineering
Ateneo de Manila University
Loyala Heights, 1108 Quezon City, Philippines

Citation
Rojas NRL. 2002. Bioinformatics: Genomics and beyond (A user's perspective). Transactions NAST PHL 24(2): 153-158. doi.org/10.57043/transnastphl.2002.5086

Trans. Nat. Acad. Sci. Tech. Philippines 24: 153-158 (2002).
ISSN 0115-8848

BIOINFORMATICS: GENOMICS AND BEYOND (A USER'S PERSPECTIVE)

NINA ROSARIO L. ROJAS
Department of Chemistry, School of Science & Engineering
Ateneo de Manila University
Loyala Heights, 1108 Quezon City, Philippines
http://www.chem.admu.edu.ph/~nina/

ABSTRACT

Bioinformatics tools play key roles in handling biological information at every stage of the research process, from formulating hypotheses, gathering, storing, retrieving and analyzing data, to publication, peer review, and validation of scientific information. Genomics has received the greatest amount of attention. However, scientists have also turned their attention to gene expression through "transcriptomics," or the study of genes that are transcribed into RNA under particular conditions. Other scientists are deep into "proteomics," or studying protein structure and function, as well as predicting gene function at the protein level. In addition, some researchers are addressing gene function by mapping out interactions between gene products, leading to what some people call the study of the "interactome." Bioinformatics, together with more efficient wet laboratory tools, have made that explosion possible. Both would be needed to understand more fully the workings of the organism living in its environment. Finally, bioinformatics is an area of opportunity for science in the Philippines.

Keywords: bioinformatics, genomics, proteomics, transcriptomics, computational tools

INTRODUCTION

Bioinformatics is the field of science that uses computational methods to answer biological questions: draw out ideas from biological databases, derive relationships among them, ask new questions and predict some answers which can tested further.

The use of computational methods is not new to biology. Computers had long been used for data analysis and model building, and the Internet was used to distribute information among the educational and research institutions that were among the first civilian members of the network. What dramatically changed the computing landscape in the 1990s were three major factors: (1) The Internet grew into a network that connected not just academic institutions but commercial entities, independent organizations, and individuals all over the globe; (2) the Internet became easy to navigate through the development of the World Wide Web and the browsers and tools that gave access to it, allowing users from all over access to computational data and tools in central servers; and (3) computing power grew to the point that ordinary personal computers could do many problems that used to require expensive workstations.

The computational developments are just one side of the story. Recognition of the value of bioinformatics as a field in itself has come about largely because of the needs of the genome projects as well as of the "post-genome" era. The sheer amount of data generated by genomics created demands not just for large-scale experimental work but also for large-scale computation. Biology had entered the world of "Big Science," just as the physicists had their supercolliders and other megaprojects. The human genome milestone caught the public imagination, such that the completion of the working drafts of the human genome was announced by the heads of state of the United States and the United Kingdom in 2000. The results of both the public and the privately funded projects were published in February 2001 in the two top science journals in the world. The post-genome era has begun.

COMPUTATIONAL NEEDS IN THE POST-GENOME ERA

What makes genomics "Big Science"? Once can approach this by asking why genomics then, rather than genetics? The genome is the complete set of genetic information of an organism, rather than just individual genes. The sheer scale of the human genome makes it a major feat for humanity to read it: about two meters of DNA per cell, or about 6 billion units of the DNA alphabet multiplied by two for maternal and paternal copies. Genomics and other "-omics" projects require parallel or fast or both parallel and fast methods of qualitative and quantitative analysis to acquire, process, store, retrieve, and analyze the huge amount of experimental information. Thus, "-omics" require "-informatics."

This need for efficient computational methods is not unique to the life sciences. As the September 14, 2001 cover of *Science* notes: "From a specialized workhorse serving the scientific elite, the computer has matured into a ubiquitous all-purpose tool for probing the universe and communicating the results."

The computational needs are apparent considering that the answers to these questions are buried in large, complex data sets (Baxevanis, 2001). The key computational issues revolve around (1) processing experimental data; (2) storage and retrieval of data, as well as data mining; (3) analysis of data for similarities and differences; (4) modeling, predicting, and discovering relationships among different pieces of information.

Although the genome project is occasionally seen in the public imagination to offer answers to everything, knowing the DNA sequence of our genome does not really give us genetic information. The individual genes would first have to be found and their functions assigned. Much of the continuing work of the genome projects centers on these problems. Yet genes only offer information on the potential toolkit of an organism, and the organisms' use of these tools depend on the environment. The question of gene function eventually becomes a question of which genes are expressed when, and what the expressed gene product (RNA or protein) does. This is based on the so-called "central dogma" of molecular biology, where the information encoded in DNA is transcribed to RNA, and translated into protein.

The challenge then is that genes cannot be studied in isolation from the environment that turns genes on and off. Neither can the effects of genes be dissociated from the environment in which the gene products act. Thus, RNA and proteins are more important, yet more difficult, to study than the DNA genome because their levels vary in response to internal and external cues and their sequence and structure are modified to fine tune their function.

EXPERIMENTAL DATA FROM POST-GENOMIC APPROACHES

Based on the so-called central dogma, scientists have turned their attention to gene expression through "transcriptomics," or the study of genes that are transcribed into RNA under particular conditions. DNA microarray technology is the major approach used in this area. Other scientists are deep into "proteomics," or studying the protein profile, protein structure and function, as well as predicting gene function through comparison with those of known proteins. In addition, some researchers are addressing gene function by mapping out interactions between gene products, leading to what some people call the study of the "interactome." Others would push "glycomics", or the study of polysaccharides, which play key roles in cellular recognition and signaling. Still other researchers talk about the "metabolome," or the small molecule metabolites of the organism, thus linking biochemistry and molecular biology with natural products chemistry.

Why all the additional "-omes"? The central dogma sounds simple: the information flows from DNA to RNA to proteins to other molecules which then interact and coordinate to make up the organism. However, the information flow is not straightforward despite the cracking of the genetic code. For example, messenger RNA is edited before being translated to protein; proteins are usually edited and modified after translation; proteins are also often conjugated with other moieties such as sugars; proteins make possible a large variety of other molecules in response to environmental signals. Finally, the organism is more complex than the sum of its parts, as its organization, or the relationships among its parts, is the key to its success. The recent discovery of RNA interference (RNAi) as a control mechanism for gene expression is just an example of how the simple view of DNA → RNA → Protein is not enough.

Hence, genomics is not enough. Experimentalists and bioinformaticists everywhere are generating other types of data to better understand living systems in their complexity. However, it is not just the "-omics" researchers who benefit. By using this treasure trove of information, life scientists who are interested in asking detailed questions about individual genes, proteins, other biological compounds, and their interactions can refine their hypotheses, design better experiments, and analyze their results in a more comprehensive manner than ever before. In addition, questions on a larger scale, such as phylogeny, metabolic pathways, and interactions between the organisms can begin to be answered at the molecular level. This can propel scientific work not just in molecular genetics but also in biochemistry, natural products, taxonomy, biodiversity, population studies, drug discovery, agriculture, medicine and other fields.

PERSPECTIVES FOR THE PHILIPPINES

How then can we make use of "-omics" and "bioinformatics" in the Philippines? The approach to bioinformatics needs to take place at several levels. The key process is not so much the hardware, the software, or the experiments per se, but what Bienvenido F. Nebres, S.J. calls "building a school of thought."

First, users of biological information, such as students and researchers in the life sciences, must learn to navigate their way among the databases and tools. They must become comfortable in using these public domain databases on genes and proteins more and more in much the same way that they check published literature and scientific indices in the course of their studies. To that end, we have started integrating the basic use of these databases into the curriculum of students by using a combination of the available materials and homegrown tutorials that introduce some key areas. Several institutions, including ours, have started offering specialized bioinformatics classes as well.

Second, we need to build an interdisciplinary community to move from simply using bioinformatics tools to adapting and modifying them and then to developing our own tools. It means building a community of researchers out of the life scientists who ask the biological questions or are generating the biological data and the computational scientists who can develop the bioinformatics tools. In our experience, this is not necessarily easy. It means sharing our different languages, our different problems, our different tools, and our different approaches. Although it is possible for a single person to be competent in the life science lab bench as well as in computation, in practice the learning curve can be steep, especially when tackling complex biological problems or challenging computational tasks. At this level, dialogue is extremely important in understanding the strengths and limitations of our own fields and approaches. One way to develop closer links among the different fields is for faculty to collaborate not just in research but also in teaching bioinformatics to groups of students from different disciplines, who can then work together on projects that go beyond simply using existing tools and resources. This is what John Paul Vergara and I have tried to do with our classes at our institution.

Third, we need to build a cross-institutional, national community. Bioinformatics research requires talent and computer hardware and software, and does not need too much laboratory bench equipment. These make bioinformatics an area of opportunity for the Philippines. Three key strategies were suggested by Richard Bruskiewich in his address during the Philippine Council for Health Research and Development Anniversary Symposium in 2002: (1) Focus upon developing and bringing together a critical mass of the necessary expertise; (2) develop core national infrastructure in bioinformatics; and (3) develop an entrepreneurial mindset. Let us have a closer look at the first two strategies.

The different classes, discussions, forums, and symposia, including those sponsored by the National Academy of Science and Technology, help advance the first point. More can be done, but we are making progress. The second part of the strategy is moving through the development of the computation infrastructure of the country through projects such as PREGINET of the Advanced Science and Technology Institute (ASTI). One necessary aspect of the infrastructure is the development of a readily searchable online archive of the abstracts of scientific literature in the Philippines, similar to PubMed. There are a few archives that are already online, but these are not yet very comprehensive. Publications in the international literature are relatively easy to find and access through PubMed, but access is more limited for papers published in local journals, conference proceedings, theses and dissertations, and scientific reports from local sources. Yet these are the works that we need to access, even in abstract form, in order to increase collaboration, reduce redundancy, and forge new directions in areas that are most relevant to the country's research agenda. In addition, these sources provide a fuller, more current picture of current RP research than the works that are already in the international journals indexed in PubMed. Improving access to archives of our literature would allow us to better cite and evaluate each other's work, thus improving the overall impact of individual research on the shape of RP science.

In summary, bioinformatics is not really a new field, but a natural and happy product of the marriage of developments in both the capacity to generate information in the laboratory and the capability of computer systems and computational methods. In the Philippines, we can tap into bioinformatics to use the genomic and post-genomic data explosion in our own laboratories; we can also contribute to it through the data analysis and computational talent that we can harness. The important thing is for life scientists and computational scientists to continue asking questions within our own context and developing methods to answer them.

SELECTED INTERNET RESOURCES

Databases and Tools to Start

The National Center for Biotechnology Information. Access to a vast, interlinked resource of abstracts and literature (PubMed), nucleotide sequences (GenBank), gene maps and genomes, protein sequences, protein structures, taxonomic data, reference materials, analysis tools such as BLAST, and the Online Mendelian Inheritance in Man (OMIM) information site. This site is an excellent site for gene-based information retrieval. <*http://www.ncbi.nlm.nih.gov/*>

The Swiss Bioinformatics Institute. Access to a suite of resources that are especially strong in protein research, including a number of tools that bridge the gap between the lab bench and the databases. The institute is also one of the key places for proteomics research. Home to the curated SwissPROT database. <*http://www.expasy.ch/*>

The Research Collaboratory for Structural Bioinformatics. Access to the Protein Data Bank (PDB), which stores data on three-dimensional structures of biological molecules. <*http://www.rcsb.org/pdb/*>

The Protein Explorer Site by Eric Martz and colleagues. Tools and tutorials for molecular visualization. Especially good for beginners and teachers. <*http://www.proteinexplorer.org*>

Tutorials Developed in Our Laboratory

Quiambao PRM and NRL Rojas. 2000. Protein structural analysis using Internet tools: A study of peroxidases. In: Proceedings of the 16th Philippine Chemistry Congress, May 22-25, 2000, Manila, Philippines. <*http://www.chem.admu.edu.ph/~nina/rosby/*>

Go, AT and NRL. Rojas. 2001. A bioinformatics toolkit for proteome analysis. In: Proceedings of the 17th Philippine Chemistry Congress, May 23-25, 2001, Cagayan de Oro, Philippines. <*http://www.chem.admu.edu.ph/~abi/*>

REFERENCES

Baxevanis AD. and BFF Ouellette, Eds. 2001. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. 2nd Edition. Wiley-Interscience, New York.

Lesk AM. 2002. Introduction to Bioinformatics. Oxford.