

Genomic empowerment: the importance of public databases

doi:10.1038/ng1186

Over the past twenty five years, a mere sliver of recorded time, the world of biology — and indeed the world in general — has been transformed by the technical tools of a field now known as genomics. These new methods have had at least two kinds of effects. First, they have allowed scientists to generate extraordinarily useful information, including the nucleotide-by-nucleotide description of the genetic blueprint of many of the organisms we care about most—many infectious pathogens; useful experimental organisms such as mice, the round worm, the fruitfly, and two kinds of yeast; and human beings. Second, they have changed the way science is done: the amount of factual knowledge has expanded so precipitously that all modern biologists using genomic methods have become dependent on computer science to store, organize, search, manipulate and retrieve the new information.

Thus biology has been revolutionized by genomic information and by the methods that permit useful access to it. Equally importantly, these revolutionary changes have been disseminated throughout the scientific community, and spread to other interested parties, because many of those who practice genomics have made a concerted effort to ensure that access is simplified for all, including those who have not been deeply schooled in the information sciences. The goal of providing genomic information widely has also inevitably attracted the interests of those in the commercial sector, and privately developed versions of various genomes are also now available, albeit for a licensing fee.

The operative principle most prominently involved in transmitting the fruits of genomics—the one that has captured the imagination of the public and served as a standard for the sharing of results and methods more generally in modern biology—has been open access. Funding by public and philanthropic organizations, such as the U.S. National Institutes of Health, the U.S. Department of Energy, the Wellcome Trust in Britain, and many other organizations, has made this altruistic behavior possible and has fostered the idea that genomic information about biological species should be available to all. (Such information about individual human beings is, of course, an entirely different matter and should be protected by privacy rules.) The attitude of open access to new biological knowledge has also been embodied in the databases of the International Nucleotide Sequence Database Collaboration, comprising the DNA DataBank of Japan, the European Molecular Biology Laboratory, and GenBank at the US National Library of Medicine. The same focus on open access is exemplified by PubMed (operated by the NLM), other gateways to the scientific literature, and the assemblies of genomic sequence now found at the several Web portals described in this guide.

The Human Genome Project (HGP), which has supported the public genome sequencing effort, has been the mainstay of the effort to make genomes accessible to the entire community of scientists and all citizens. This effort has, in fact, been quite naturally extended to instruct the public about many themes in modern biological science. This has occurred in part because the human genome itself has been such an exciting concept for the public; in part because genomes are natural entry points for

teaching many of the principles of biological design, including evolution, gene organization and expression, organismal development, and disease; and in part because those who work on genomes have been tireless in attempts to explain the meaning of genes to an eager public. Endless metaphors, artistic creations, lively journalism, monographs about social and ethical implications, televised lectures from the White House, and many other cultural happenings have been among the manifestations of this fascination. In this way, the HGP has had a strong hand in raising the public's awareness of new ideas in biology and of the powerful implications of genomics in medicine, law and other societal institutions.

Some of these cultural effects come as much from the behavioral aspects of the HGP as from the genomic sequences themselves. The sharing of new information, even before its assembly into publishable form, has spurred efforts to share other kinds of research tools and has encouraged the notion of making the scientific literature freely accessible through the Internet. The contribution of scientists in many countries to the sequencing of many genomes, including the human genome, has inspired efforts to develop gene-based sciences—from basic genomics to biotechnology—throughout the world, including the poorest developing nations. Indeed, the World Health Organization, the United Nations, and the World Bank have all contributed recently to the growth of the ideas that science is both possible and valuable in all economies and that science can be a means to help unify the world's population under a banner of enlightenment, demonstrating a virtue of globalization.

From this perspective, the availability of the sequences of many genomes through the Internet is a liberating notion, making extraordinary amounts of essential information freely accessible to anyone with a desktop computer and a link to the World Wide Web. But the information itself is not enough to allow efficient use. Interested people who reside outside the centers for studying genomes need to be told where best to view the information in a form suitable for their purposes and how to take advantage of the software that has been provided for retrieval and analysis.

The manual before us now offers such help to those who might otherwise have had trouble in attempting to use the products of genomics. Furthermore, the advice is offered in that spirit of altruism that has come to characterize the public world of genomics. The information is provided in a highly inviting and understandable format by casting it in the form of answers to the questions most commonly posed when approaching big genomes. The information, made freely available on the World Wide Web, has been assembled by some of the best minds in the HGP, who have generously given their time and intellect to encourage widespread use of the great bounty that has been created over the past two decades.

In other words, the guide to use of genomes provided here is simply another indication that the HGP should take great pride in much more than the sequencing of genomes.

Harold Varmus
Memorial Sloan-Kettering Cancer Center