

Editorial

Evolutionary Annotation of the Genome

Areas of biomedical research are being referred to as “genomic annotation.” This follows the announcement earlier this year of the near completion of the human genome sequence. The metaphor is usually lexicographical: The genome is a dictionary filled with words whose meanings we are challenged to discover. Organisms are books, which we will be able to read once we understand the words. There is talk, in the context of model species, of “Rosetta Stones” and “Chaucerian English.”

This is a poor metaphor. We (I mean non-Russian speakers) can never hope to understand *The Brothers Karamazov* equipped only with an annotated English-Russian dictionary. We need some knowledge of Western philosophy, orthodox Christian theology, European history, 19th century Russian political and social organizations—knowledge of the life and times of Mikhailovich Dostoyevski. To understand an organism, we need to know about its ecology, its physiology, and its cell and developmental biology. The genome and its immediate products are critically important, but they are not enough. This is not news. The challenge of understanding genotype–phenotype relationships is an old one. The outcome of genome projects is that we now have the prospect of meeting this challenge on a grand scale.

The annotation metaphor is, in fact, not derived from lexicography; it comes from computer science. A software engineer annotates her code to explain to potential users how parts of it work or why they have been written in a particular way. Explanation is provided about the interaction of elements of code in the context of a functioning program, rather than about the immediate meanings of individual words. This is a more apt conceptual approach to understanding organisms. It also focuses attention on the engineer.

In biology the process of organic evolution replaces the engineer, and genomes come without annotation. If we want to understand how genomes function in the context of organisms, we need to look at the evolutionary processes that gave rise to them; molecular evolutionary analysis becomes central to postgenomics biology.

In this issue of *Molecular Biology and Evolution*, we publish a perspectives article by David Pollock and colleagues that describes the Vertebrate Mitochondrial Genome Project as a pilot for large-scale evolutionary genomics. The project aims to connect systematists (and their supplies of tissues and taxonomic knowledge); genomic researchers, who can efficiently clone and sequence on a large scale; computational biologists and

bioinformaticians, who can develop and implement approaches to data analysis; and functional and structural biologists, who can incorporate the analysis into an understanding of cellular function and physiology.

SMBE has an important role to play in promoting, fostering, and encouraging this area of investigation through the development of a coordinated, community-wide approach to the field of evolutionary genomics. David Pollock and I will shortly launch an Evolutionary Genomics and Sequence Biodiversity Web Site on behalf of SMBE.

This SMBE site will provide a forum where research groups can pool efforts and jointly promote their work. It will help organize the large-scale production and interpretation of primary sequence information and build a collection of bioinformatics tools for manipulating, organizing, and analyzing information in an evolutionary context. The site will provide the organizational structure for a coordinated international effort to develop an evolution-based understanding of genome organization, structure, and function.

The immediate goals of the site are:

- to provide access and organization to phylogeny- and evolution-based bioinformatic tools that facilitate large-scale sequence biodiversity data sets (both long multi-gene sequences and large numbers of taxa);
- to coordinate programming efforts and software evaluation and distribution;
- to provide easier comparative analysis of classes of functional groups (e.g., secondary structure, transmembrane regions, gene boundaries);
- to facilitate the testing, publication, and distribution of new analytical solutions from different laboratories;
- to promote a communitywide, open-source evolutionary genomics approach, fostering links to other genome research efforts;
- to promote links between sequence information and the sample sources in an effort to promote conservation of the original sources of biodiversity and to promote efforts by museums, zoological parks, and related organizations to acquire and curate the source material;
- to promote an increase in the scale of sequence and genome biodiversity projects;
- to advance the field by promoting projects in the form of press releases and other publicity efforts.

We welcome the participation of other SMBE members in helping to establish a framework for evolutionary genome annotation.

Simon Easta
Editor