Wojciech Makalowski

Pennsylvania State University, Department of Biology
and Institute of Molecular Evolutionary Genetics,
Pennsylvania, USA

CV

Wojciech Makalowski graduated from Adam Mickiewicz University in Poznan, Poland with a master degree in philosophy of science and a Ph.D. in molecular biology. Very early in his scientific carrier he "discovered" that in biology nothing makes sense except in the light of evolution. With the strong molecular biology background, the molecular evolution was an obvious choice to pursue his interest. As computational biologist he went through a series of system reincarnations. The first useful for biologists program he wrote in BASIC for a ZX Spectrum machine with impressive 48 kb RAM and a cassette tape recorder as a storage device. Later, still in Poland, he moved on "real" machine - IBM PC clone and PASCAL as programming language. After a short postdoc in Montreal, Canada, he joined National Center for Biotechnology Information (NCBI) in Bethesda, Maryland, USA in 1994 where he stayed until last summer working in the Mark Boguski's group where he joined a horde of UNIX enthusiasts. Recently, he joined the Institute of Molecular Evolutionary Genetics and the Department of Biology at the Pennsylvania State University in State College, USA. He is a member of editorial board of the Gene and the Genome Research journals. His research is focused on eukaryotic genomes evolution especially influence of transposable elements on their host genome - the concept of the genomic scrap yard and comparative genomics of Eucaryotes. The latter involves development of methods for the orthologous and paralogous sequences discrimination, as well as visualization tools for large scale sequence analyses.

CONTACT INFORMATION

Wojciech Makalowski, Ph.D.
National Institutes of Health
National Center for Biotechnology Information
45 Center Drive, MSC 6510, Bldg. 45, Rm. 6As.47A
Bethesda, MD 20892-6510, USA
phone: (301)435-5989; fax: (301)480-2918
makalow@ncbi.nlm.nih.gov

Computational Genomics - Promises and Challenges

Genome science and technology has brought us to the brink of being able to describe the genetic blueprint and molecular evolutionary history of the human species. Now we can read the genomes at a frightening speed - a several-million-base bacterial genome in several days is not a problem for one of the major sequencing centers, and a billion-base eukaryotic genome can be done in several months. But reading a text and understanding it are two different things. So, would the genomic community pass the genome comprehension test? The answer to this question is vital to the whole genomic enterprise, and this is where computational analysis of genomes takes a central stage. I will discuss major goals, problems, and challenges of computational genomics.

The prevailing view of genomes is gene-centric and therefore the major effort of genomic community is to accurately annotate those important genomic features. Most of a prokaryotic genome encodes for the proteins. In contrary, protein-coding sequences in eukaryotic genomes occupy only a minute fraction of a genome. Additionally, the information about a single protein is interrupted but sometimes very long non-coding sequences. Hence, gene prediction in prokaryotic and eukaryotic genomes are of different nature. Although, the gene prediction problem in prokaryotic genomes can be considered as solved, the gene prediction in eukaryotic genomes is one of the biggest bioinformatics challenges. The extent of the problem is such that it is not likely that can be solved by computational methods only. With the human genome in hand we look toward other mammalian genomes for help. We will need several genomes of different evolutionary distances along with the information on expressed sequences (ESTs, full-length mRNAs, etc.) to accurately annotate our genome. Once the predicted proteome for a given genome has been defined, the major task is to analyze and interpret it. Again, comparative genomic is a major force here. Database searches became everyday task of most molecular biologist. I will discuss current efforts to improve homology search sensitivity including domain profiles based approaches.

Although most of an eukaryotic genome does not encode proteins, these sequences are not useless. They carry a biological treasure hidden from most current methods. Only by comparing syntenic regions of several organisms we can detect regions which are subject of stabilizing evolution which in turns indicates functional importance of a revealed region.

The problems outlined above are just the tip of the iceberg that bioinformatics has to deal with. The comparative genomics is a central approach in most of the cases. Ergo, the most important achievement of the Human Genome Project is fact that it has spawned sequencing of other genomes from different branches of the tree of life, including multiple species of bacteria, archaea, fungi, plants, and animals. Genome comparison is crucial in interpreting of any particular genome.