

Systems Biology Research Symposium

Oral Presentation Session

Grand Ballroom
Tuesday, June 5th
7:00-8:30pm

BioBIKE: Providing Biologists without Programming Experience Creative Control over the Analysis of Bioinformation

Jeff Elhai¹, Arnaud Taton², JP Massar³, and Jeff Shrager⁴

¹Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA, USA, ²Section of Molecular Biology, University of California, San Diego, CA, USA, ³Berkeley, CA, USA, ⁴CollabRx, Inc., Palo Alto and Symbolic Systems Program, Stanford University, Stanford, CA, USA

Presenter's email address: elhaij@vcu.edu

The recent availability of vast amounts of data in the forms of genomic sequences and mass experimental data sets has been the engine of rapid progress in many areas of biological research. But progress has tended to be broad rather than deep, because few researchers are able to build specialized computational tools to look at information in fundamentally new ways. Basic new insights occur most often to those minds primed by experience and fed by novel encounters with data not already filtered by existing models. Mere collaboration between one person with the experience and another with the data will not do. Agility in today's information-rich world demands the ability to work creatively with the large data sets and therefore to program the computer.

Researchers generally have not expended the significant effort required to learn a conventional programming language. If biologists will not come to programming, then programming must go to biologists.

We have developed a web-based resource, BioBIKE (Biological Integrated Knowledge/programming Environment), that combines available genomic, metabolic, and experimental knowledge with tools available through a graphical interface. Most importantly, it provides a means of inventing new tools through a language that incorporates the concepts of molecular biology.

While the presentation within BioBIKE is generally understandable to biologists, it is more difficult for naïve users to find the desired functionality. We are working towards a natural language interface that engages the user in a conversation leading either to a computational expression sought by the user or to an appropriate help screen. We are also combining BioBIKE with the SEED, a community annotation resource, to integrate the analysis and annotation of genomes by means accessible to biologists without programming experience.

Key words: visual programming, human-machine interface, genome analysis, natural language