

Web-based tools, called 'Galaxy,' simplify genomic analysis

February 23 2010, By David Pacchioli

With tremendous advances in DNA sequencing and the advent of microarray technology in the 1990s, biology embarked on a new age of discovery. Researchers suddenly had access to unprecedented amounts of data -- and faced unprecedented complexity in its analysis.

Necessity sparked the rise of a whole new field: the hybrid of biology and computer science now known as [bioinformatics](#). But as sequencing technologies continue to evolve more and more rapidly, the challenge has grown more and more acute.

"Biology is in a state of shock," says Anton Nekrutenko, assistant professor of biochemistry and molecular biology at Penn State. "What we have is biochemistry and biology labs that are generating mountains of data, and then they say, 'What do we do now?'"

"[Computational biologists](#) write the programs they need to solve their own problems," Nekrutenko said, "but they are generally not interested in providing interfaces for experimental biologists."

That's where Galaxy comes in. Developed by Nekrutenko and others at Penn State, along with James Taylor at Emory University, Galaxy is a Web-based framework that pulls together a variety of tools that allow for easy retrieval and analysis of large amounts of data, simplifying the process of genomic analysis. As described in one of the team's early papers, Galaxy "combines the power of existing genome annotation databases with a simple Web portal to enable users to search remote resources, combine data from independent queries, and visualize the

results.”

“Essentially we are providing a unified interface to many different tools,” Nekrutenko explains. As a trade review puts it, Galaxy “amplifies the strengths of existing resources.”

The response has been gratifying, to say the least. “Since last year the project has really taken on legs,” Nekrutenko said. The Galaxy website at Penn State now has 10,000 registered users, and many more who are not registered. It runs 4-5,000 analyses daily. “It’s also available as software, so that people can download it and to run it anywhere, on their own hardware. We encourage this, in fact, because there’s a limit to how much data our computers can handle.

“Our goal is proliferation,” Nekrutenko said, “and right now we don’t have much competition. We are really the only genomic solution. We allow biologists to do various very complicated analyses quite easily. And we have all sorts of cool features,” including an automated workflow management tool and a host of short video tutorials. “There’s even an iPhone app so you can check your analysis as it’s running,” he said.

As with most of the software in this rapidly evolving field, Galaxy is completely open source. “That’s how biology works these days,” Nekrutenko said. “There are commercial solutions, but it’s a waste of money because the technology changes every two weeks.”

He and his collaborators continue to work on improvements. One of their current aims is to make computational analyses transparent and reproducible, a basic tenet of experimental research. Nekrutenko points to one of his own papers, recently published in the journal *Genome Research*. With the aid of Galaxy, every stage of the analysis that he and his co-authors conducted for their study is published as supplementary

data, alongside the online version of the article.

"We envision being able to do this with other journals," he said. "At every step, an interested reader will be able to go through the data."

He said the pace of change keeps things interesting.

"There are emerging technologies that will produce 100 times more data than the so-called next-generation sequencing. We're already at next-next-generation sequencing. It's reaching the point where storage becomes an issue, never mind analysis."

It's exciting to be in the middle of such ferment, and also stressful, "but we have a very good team assembled and a lot of momentum. We have had generous early support from the Huck Institutes at Penn State, and we are now well-funded by NSF and NIH. The funding agencies have finally recognized that they need to pay not only for data generation, but also for data management," Nekrutenko said. "I think we're in a really good place."

More information: main.g2.bx.psu.edu/

Provided by Pennsylvania State University

Citation: Web-based tools, called 'Galaxy,' simplify genomic analysis (2010, February 23)
retrieved 23 April 2024 from

<https://phys.org/news/2010-02-web-based-tools-galaxy-genomic-analysis.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.