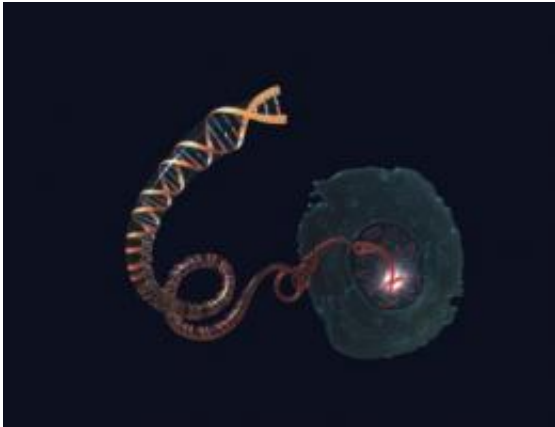


# Galaxy DNA-analysis software is now available 'in the cloud'

November 8 2011

---



Galaxy -- an open-source, web-based platform for data-intensive biomedical and genetic research -- is now available as a "cloud computing" resource. The new technology, developed by a team of researchers from Penn State University and Emory University, will help scientists and biomedical researchers to harness such tools as DNA-sequencing and analysis software, as well as storage capacity for large quantities of scientific data. More information is online at

<http://science.psu.edu/news-and-events/2011-news/Nekrutenko11-2011> Credit:

National Institutes of Health

Galaxy -- an open-source, web-based platform for data-intensive biomedical and genetic research -- is now available as a "cloud computing" resource. A team of researchers including Anton Nekrutenko, an associate professor of biochemistry and molecular biology at Penn State University; Kateryna Makova, an associate

professor of biology at Penn State; and James Taylor from Emory University, developed the new technology, which will help scientists and biomedical researchers to harness such tools as DNA-sequencing and analysis software, as well as storage capacity for large quantities of scientific data. Details of the development will be published as a letter in the journal *Nature Biotechnology*.

Earlier papers by Nekrutenko and co-authors describing the technology and its uses are published in the journals *Genome Research* and *Genome Biology*.

Nekrutenko said that he and his team first developed the [Galaxy computing system](http://galaxyproject.org) (<http://galaxyproject.org>) in 2005 because "biology is in a state of shock. Biochemistry and biology labs generate mountains of data, and then scientists wonder, 'What do we do now? How do we analyze all these data?'" Galaxy, which was developed at Penn State and continues to use the University's servers for its computing power, solves many of the problems that researchers encounter by pulling 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 in the journal *Genome Research*, 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." Galaxy also allows other researchers to be able to review the steps that have been taken, for example, in the analysis of a string of [genetic code](#). "Galaxy offers scientific transparency -- the option of creating a public report of analyses. So, after a paper has been published, scientists in other labs can do studies in order to reproduce the results described," Nekrutenko said.

Now, Nekrutenko's team has taken Galaxy to the next level by developing an "in the cloud" option using, for example, the popular Amazon Web Services cloud. "A cloud is basically a network of

powerful computers that can be accessed remotely without the need to worry about heating, cooling, and system administration. Such a system allows users, no matter where they are in the world, to shift the workload of software storage, data storage, and hardware infrastructure to this remote location of networked computers," Nekrutenko explained.

"Rather than run Galaxy on one's own computer or use Penn State's servers to access Galaxy, now a researcher can harness the power of the cloud, which allows almost unlimited computing power." As a case study, the authors report on recent research published in *Genome Biology* in which scientists, with the help of Ian Paul, a professor of pediatrics at Penn State's Hershey Medical Center, analyzed DNA from nine individuals across three families using Galaxy Cloud. Thanks to the enormous computing power of the platform, the researchers were able to identify four heteroplasmic sites -- variations in mitochondria, the part of the genome passed exclusively from mother to child.

"Galaxy Cloud offers many advantages other than the obvious ones, such as [computing power](#) for large amounts of data and the ability for a scientist without much computer training to use DNA-analysis tools that might not otherwise be accessible," Nekrutenko said. "For example, researchers need not invest in expensive computer infrastructure to be able to perform data-intensive, sophisticated scientific analyses."

Yet another advantage of Galaxy Cloud is its data-storage capacity. Using the Amazon Web Services cloud, researchers have the option of storing vast amounts of data in a secure location. "There are emerging technologies that will produce 100 times more data than existing 'next-generation' DNA sequencing, which already has reached the point where even more storage becomes an issue, not to mention analysis," Nekrutenko said.

Provided by Pennsylvania State University

Citation: Galaxy DNA-analysis software is now available 'in the cloud' (2011, November 8)  
retrieved 23 April 2024 from  
<https://phys.org/news/2011-11-galaxy-dna-analysis-software-cloud.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.