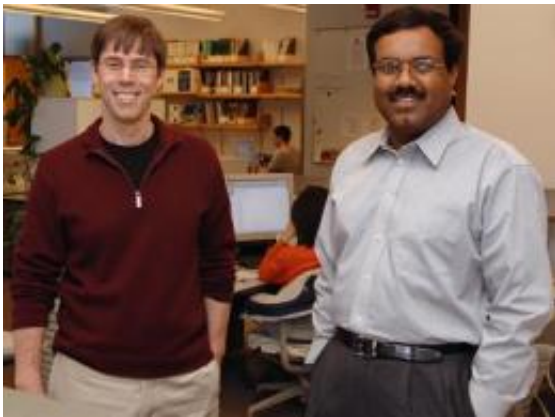


# Iowa State researchers developing 'BIGDATA' toolbox to help genome researchers

October 3 2012

---



Iowa State University's Patrick Schnable, left, and Srinivas Aluru are developing a toolbox to help life sciences researchers analyze all of the data produced by today's DNA sequencing instruments. Credit: Photo by Bob Elbert/Iowa State University

Today's life scientists are producing genomes galore.

But there's a problem: The latest DNA sequencing instruments are burying researchers in trillions of bytes of data and overwhelming existing tools in biological computing. It doesn't help that there's a variety of sequencing instruments feeding a diverse set of applications.

Iowa State University's Srinivas Aluru is leading a research team that's

developing a set of solutions using high performance computing. The researchers want to develop core techniques, parallel algorithms and software libraries to help researchers adapt parallel computing techniques to high-throughput DNA sequencing, the next generation of [sequencing technologies](#).

Those technologies are now ubiquitous, "enabling single investigators with limited budgets to carry out what could only be accomplished by an international network of major sequencing centers just a decade ago," said Aluru, the Ross Martin Mehl and Marylyne Munas Mehl Professor of Computer Engineering at Iowa State.

"Seven years ago we were able to [sequence DNA](#) one fragment at a time," he said. "Now researchers can read up to 6 billion [DNA sequences](#) in one experiment.

"How do we address these big data issues?"

A three-year, \$2 million grant from the BIGDATA program of the National Science Foundation and the National Institutes of Health will support the search for a solution by Aluru and researchers from Iowa State, Stanford University, Virginia Tech and the University of Michigan. In addition to Aluru, the project's leaders at Iowa State are Patrick Schnable, Iowa State's Baker Professor of Agronomy and director of the centers for Plant Genomics and Carbon Capturing Crops, and Jaroslaw Zola, a former research assistant professor in electrical and computer engineering who recently moved to Rutgers University.

The majority of the grant – \$1.3 million – will support research at Iowa State. And Aluru is quick to say that none of the grant will support hardware development.

Researchers will start by identifying a large set of building blocks

frequently used in genomic studies. They'll develop the parallel algorithms and high performance implementations needed to do the necessary data analysis. And they'll wrap all of those technologies in software libraries researchers can access for help. On top of all that, they'll design a domain specific language that automatically generates computing codes for researchers.

Aluru said that should be much more effective than asking high performance computing specialists to develop parallel approaches to each and every application.

"The goal is to empower the broader community to benefit from clever parallel algorithms, highly tuned implementations and specialized high performance computing hardware, without requiring expertise in any of these," says a summary of the research project.

Aluru said the resulting software libraries will be fully open-sourced. Researchers will be free to use the libraries while developing, editing and modifying them as needed.

"We're hoping this approach can be the most cost-effective and fastest way to gain adoption in the research community," Aluru said. "We want to get everybody up to speed using high performance computing."

Provided by Iowa State University

Citation: Iowa State researchers developing 'BIGDATA' toolbox to help genome researchers (2012, October 3) retrieved 16 April 2024 from <https://phys.org/news/2012-10-iowa-state-bigdata-toolbox-genome.html>

<p>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.</p>
--