genome web

NSF Awards $1M to Support Petascale Computational Tools for Genome Rearrangement Research

November 18, 2009

The National Science Foundation has funded three research teams under a four-year $1 million project to help ensure that certain bioinformatics tools are compatible with upcoming petascale computers.

NSF awarded the grants under the American Recovery and Reinvestment Act to support the development of algorithms that will infer evolutionary relationships from genomic rearrangement events — a task that could take "centuries" to analyze on today's fastest parallel computers, according to the agency.

As a result, these algorithms will be designed to run on machines that can process more than a thousand trillion calculations per second. Currently, only two such systems are installed in the world, according to the latest "Top500" ranking of the world's fastest supercomputers — Cray's "Jaguar," located at the Department of Energy's Oak Ridge Leadership Computing Facility, which clocked in at 1.75 petaflops per second; and IBM's "Roadrunner" system at Los Alamos National Laboratory, which recorded a performance of 1.04 petaflops per second.

NSF awarded the grants to David Bader from the Georgia Institute of Technology, Jijun Tang of the University of South Carolina; and Stephen Schaeffer at Pennsylvania State University.

According to the NSF awards database, Bader received a grant for $400,000, Schaeffer received $275,000, and Tang's team received $324,968.

"Genome sequences are now available for many organisms, but making 'biological sense' of the available genomic data requires high-performance computing methods and an evolutionary perspective, whether scientists are trying to understand how genes of new functions arise, why genes are organized as they are in chromosomes, or why these arrangements are subject to change," Bader said in a statement.

According to the grant abstract, the project to develop new algorithms and high-performance software for data analysis will allow these problems to "be addressed at scale" for the first time.

Fruit fly genomes will be the "primary source of data to assess models and methods developed." This effort leverages past work on genome rearrangement analysis through parsimony and other phylogenetic algorithms.

Specifically, the starting point for the new algorithms will be Bader's <u>Genome Rearrangements Analysis under Parsimony and other Phylogenetic Algorithm, or GRAPPA</u>, an open-source package for breakpoint analysis initially released in 2000, which is able to reconstruct the evolutionary relatedness among species.

Fruit flies are good models for studying rearrangement "because the genome sizes are relatively small for animals, the mechanism that alters gene order is reasonably well understood, and the evolutionary relationships among the 12 sequenced genomes are known," Schaeffer said in a statement.

The scientists said that the fruit fly gene order diversity results can be extended for research on mammalian genomes.

The NSF program to accelerate scientific discovery and engineering through petascale simulations and analysis, called PetaApps for short, was launched in 2007 by the NSF's Office of Cyberinfrastructure.

-->
Drupal