

Public release date: 17-Nov-2009

[[Print](#) | [E-mail](#) | [Share](#)] [[Close Window](#)]



Contact: Abby Vogel

avogel@gatech.edu

404-385-3364

[Georgia Institute of Technology Research News](#)

Petascale computing tools could provide deeper insight into genomic evolution

Technological advances in high-throughput DNA sequencing have opened up the possibility of determining how living things are related by analyzing the ways in which their genes have been rearranged on chromosomes. However, inferring such evolutionary relationships from rearrangement events is computationally intensive on even the most advanced computing systems available today.

Research recently funded by the American Recovery and Reinvestment Act of 2009 aims to develop computational tools that will utilize next-generation petascale computers to understand genomic evolution. The four-year \$1 million project, supported by the National Science Foundation's PetaApps program, was awarded to a team of universities that includes the Georgia Institute of Technology, the University of South Carolina and The Pennsylvania State University.

"Genome sequences are now available for many organisms, but making biological sense of the genomic data requires high-performance computing methods and an evolutionary perspective, whether you 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," said lead investigator David A. Bader, a professor in the Computational Science and Engineering Division of Georgia Tech's College of Computing.

Even on today's fastest parallel computers, it could take centuries to analyze genome rearrangements for large, complex organisms. That is why the research team -- which also includes Jijun Tang, an associate professor in the Department of Computer Science and Engineering at the University of South Carolina; and Stephen Schaeffer, an associate professor of biology at Penn State -- is focusing on future generations of petascale machines, which will be able to process more than a thousand trillion, or 10^{15} , calculations per second. Today, most personal computers can only process a few hundred thousand calculations per second.



IMAGE: Georgia Tech computing professor David A. Bader is leading an effort to develop computational tools that will utilize next-generation petascale computers to understand genomic evolution.

[Click here for more information.](#)

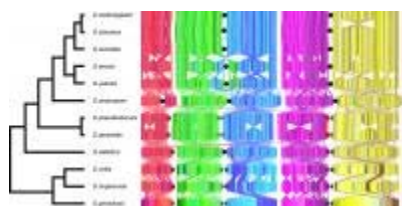


IMAGE: Microscale rearrangements of genes among 12

The researchers plan to develop new algorithms in an open-source software framework that will utilize the capabilities of parallel, petascale computing platforms to infer ancestral rearrangement events. The starting point for developing these new algorithms will be GRAPPA, an open-source code co-developed by Bader and initially released in 2000 that reconstructed the evolutionary relatedness among species.

"GRAPPA is currently the most accurate method for determining genome rearrangement, but it has only

Drosophila species. Each gene is indicated by a colored line, showing how gene order is shuffled during the evolution of these species.

[Click here for more information.](#)

been applied to small genomes with simple events because of the limitation of the algorithms and the lack of computational power," explained Bader, who is also executive director of high-performance computing at Georgia Tech.

On a dataset of a dozen bellflower genomes, the latest version of GRAPPA determined the flowers' evolutionary relatedness one billion times faster than the original implementation that did not utilize parallel processing or optimization.

The researchers will test the performance of their new algorithms by analyzing a collection of fruit fly genomes.

"Fruit flies -- formally known as *Drosophila* -- are an excellent model system for studying genome 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," said Schaeffer.

The analysis of genome rearrangements in *Drosophila* will provide a relatively simple system to understand the mechanisms that underlie gene order diversity, which can later be extended to more complex mammalian genomes, such as primates.

The researchers believe these new algorithms will make genome rearrangement analysis more reliable and efficient, while potentially revealing new evolutionary patterns. In addition, the algorithms will enable a better understanding of the mechanisms and rate of gene rearrangements in genomes, and the importance of the rearrangements in shaping the organization of genes within the genome.

"Ultimately this information can be used to identify microorganisms, develop better vaccines, and help researchers better understand the dynamics of microbial communities and biochemical pathways," added Bader.

###

This material is based upon work supported by the National Science Foundation (NSF) under Award Nos. OCI-0904461, 0904179 and 0904166. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the researchers and do not necessarily reflect the views of the NSF.

[[Print](#) | [E-mail](#) | [Share](#)] [[Close Window](#)]

