

[Log in or Register](#)

SEARCH

GO

Tuesday, March 22, 2011

[Home](#) » [News](#) » [BioInform](#)

OSC Supercomputer, Glenn, Comes in Handy for Protein Evolution Study

March 18, 2011

By [Uduak Grace Thomas](#)

Researchers at Ohio State University recently leveraged a system at the Ohio Supercomputer Center for a study that shed light on a protein family's history.

The team, led by Rebecca Lamb, an assistant professor of molecular genetics at OSU, used OSC's 9,300-core "Glenn" cluster to perform sequence alignments and generate phylogenetic trees in order to study the evolutionary history of the poly(ADP-ribose)polymerase, or PARP, protein superfamily. The researchers [published their findings](#) in *BMC Evolutionary Biology*.

"This is computationally intensive work that would have been impossible without the computer resources provided at OSC," Lamb said in a statement. "In particular, the ability to try a variety of tools that require a great deal of CPU and memory capabilities was essential for success."

PARP proteins are found in eukaryotes – animals, plants, molds, fungi, algae and protozoa — though they have been most extensively studied in mammals.

Furthermore, "PARPs have been shown to be involved in DNA damage repair, cell death pathways, transcription, and chromatin modification/remodeling," the researchers wrote. For example, a polymorphism in human PARP1 has been associated with an increased cancer risk and a decreased risk of asthma.

"We [were] interested in the fact that there are so many different types [of PARP proteins] and [that] they are spread across the eukaryotes," Lamb said. "We became interested in the evolutionary history and as we dived into it [we] found out how complex it was since the majority of functional research into this group of proteins has been done in mammals."

For the study, the researchers identified 236 PARP proteins from 77 species across five of the six eukaryotic supergroups. Lamb's team then used the Glenn cluster to perform extensive phylogenetic analyses of the identified PARP regions.

Glenn is a 9,304-core IBM Cluster 1350 system. With a maximum performance of more than 75 teraflops, it holds the No. 191 spot in the latest Top500 supercomputer ranking. The system includes AMD Opteron multi-core technologies and IBM cell processors as well as a variety of memory and processor configurations.

OSC offers several bioinformatics packages for use on Glenn, including Amber, BioPerl, Blast, Blat, ClustalW, EMBOSS, HMMer, MrBayes, ParaView, and PAUP.

For their first step, the team identified more than 300 PARP sequences using the catalytic domain that characterizes the family, called the "PARP signature."

The researchers selected sequences from the Pfam database identified as members of the PARP family and then retrieved the full sequences from the UniProt database. Additional sequences were obtained by performing Blast searches of databases

Type size: | |

SHARE

Email

Printer-friendly version

RSS Feed

1. **GenomeWeb Daily News**

Illumina Researcher Outlines Strategy for Interpreting Clinical Information in Healthy Genomes

2. **GenomeWeb Daily News**

Life Technologies Starts Trials of Sequence-based Typing System

3. **GenomeWeb Daily News**

Buybacks in Bloom as Life Science Tools, MDx Firms Deploy Cash

4. **GenomeWeb Daily News**

Exome Sequencing Studies Tackle Increasingly Complex Diseases

5. **GenomeWeb Daily News**

Genomics Key to NIDDK's Diabetes Plans

In this issue of BioInform

UCSC Genome Browser Turns to 'Data Hubs' to Avoid Transmission Bottlenecks for Large Datasets

Complete Genomics Teams with DNAnexus to Offer Clients Cloud-based Informatics for Sequence Data

Partek Opens Offices in Europe, Asia; Prepares to Launch Cloud-based NGS Software

OSC Supercomputer, Glenn, Comes in Handy for Protein Evolution Study

[Young Investigator Profile](#)[Blog](#)[Papers of Note](#)[People on the Move](#)

GenMark Diagnostics has tapped **Paul Ross** to serve as chief financial officer, a position he will assume in April. Ross currently is CFO at **Teledata**, and he formerly held the same position at **Meadco Instruments**

containing protein data from eukaryotic genomes, including resources at the Joint Genome Institute, the Broad Institute, the J. Craig Venter Institute, and the *Arabidopsis* Information Resource.

Lamb's team pared the list down to about 200 proteins by eliminating duplicates and sequences that were less than 100 amino acids long and didn't include the catalytic domain. Furthermore, the team also selected orthologs from single representative species for each group of vertebrates — for example, human for mammals and chicken for birds — and then discarded sequences from other vertebrates in these groups

"Once we had the sequences, we extracted the catalytic domain, and then we needed to align the sequences," Lamb said.

"That's where we really started using the supercomputer center a lot because this takes a lot of computational power."

The investigators selected MUSCLE to perform the alignments after comparing the results of several alignment tools offered by OSC. Lamb said that the software proved to be "the best at handling the sequences" and that it also introduced fewer gaps than other tools.

Next, they ran PhyML to generate maximum-likelihood trees based on the aligned PARP catalytic domains. Using Glenn, the team was able to test different settings for the software and compare the results to select the most optimal settings prior to creating the tree.

"It would have been extremely difficult if not impossible to run it just on our local computers," she said. "A lot of the analyses still [took a while], even [on] the supercomputer."

For instance, once they had selected the requisite software settings, it took six to seven hours to run PhyML on Glenn. However, the same analysis "might have taken days to do on our local computer," Lamb said.

The team also ran statistical tests to ensure that the trees were accurate. These tests work by randomizing the order of the alignments and generating several trees that are compared to the input tree to see the number of times the randomly generated trees match.

Based on their trees, the researchers concluded, among other things, that ancestral PARP proteins would have had different functions and activities including DNA damage response and genomic maintenance and that the diversity of the superfamily "is larger than previously documented, suggesting [that] as more eukaryotic genomes become available, this gene family will grow in both number and type."

As a next step, Lamb's team plans to explore whether PARP proteins found in some fungi play a role in their pathogenicity.

Have topics you'd like to see covered in *BioInform*? Contact the editor at [uthomas \[at\] genomeweb \[.\] com](mailto:uthomas@genomeweb.com).

Related Stories

[Three New Life Science Machines Join Top500 Supercomputer Ranking, Bringing Total to Seven 🍌](#)
November 19, 2010 / [BioInform](#)

and at **Power-One**.

Exact Sciences has named **John Krayacich** to be senior VP of sales and marketing. Krayacich recently was president and CEO of **Ambrose Pharmaceuticals** and **Marinus Pharmaceuticals**, where he worked in strategic and marketing planning and business development and financial management. He also was VP and therapeutic area head at Novartis Pharmaceutical and VP and therapeutic group leader at Pfizer Neurosciences.

Qiagen has proposed for election to its supervisory board **Elizabeth Tallet** and **Vera Kallmeyer**. Tallet has been president and CEO of **Transcell Technologies**, president of **Centocor Pharmaceuticals**, and director of worldwide strategic planning for **Warner Lambert**. Kallmeyer is a professor in the Department of Neurosurgery at the **Stanford University School of Medicine**, and she formerly was VP of corporate development at **Aviron**. Qiagen's shareholders will vote on their election at the firm's annual general meeting, which is scheduled for June 30.

Upcoming Events