



# Cyberinfrastructure Resources for Genomics Research

by Barbara Hallock, Carrie Ganote, Melissa Pespeni, *Indiana University*  
{bahallock;cganote}@iu.edu; mpespeni@indiana.edu

## ABSTRACT

New DNA sequencing technologies are generating more sequence data, faster, and cheaper. But there is a catch: the sequences are shorter and the nucleotide identification has higher error rates, meaning that the computational challenge of assembling a full genome from sequence data is also greater. In this poster, we examine cyberinfrastructure resources available to researchers undertaking genomics work, and present a case study that illustrates how one lab is currently making use of these resources.

## NATIONAL CENTER FOR GENOME ANALYSIS SUPPORT (NCGAS)

The mission of the National Center for Genome Analysis Support is to enable the biological research community of the US to analyze, understand, and make use of the vast amount of genomic information now available. NCGAS focuses particularly on transcriptome- and genome-level assembly, phylogenetics, metagenomics/transcriptomics and community genomics. To that end, NCGAS provides consulting, technical support, cyberinfrastructure resources, and other services to those with a need for genomics expertise in their experiments. NCGAS is funded by the NSF to provide these services, so depending on funding sources and project needs, cost to the researcher may be as little as a mention in the acknowledgments of any publications completed with their support.

## COMPUTATIONAL RESOURCES

In addition to support, NCGAS and XSEDE provide a number of supercomputing systems that are configured to handle the kinds of large-memory jobs required to perform genomics research. These include:

- **Mason**, a large memory supercomputer cluster at IU
- **Rockhopper**, a “cluster on demand” cloud resource where researchers can purchase time on a real supercomputer cluster owned by Penguin Computing Inc.
- **Stampede** – the largest supercomputer accessible as part of XSEDE, operated by the Texas Advanced Computing Center
- **Gordon** – a supercomputer using flash memory to support large data problems, operated by the San Diego Supercomputer Center
- **Blacklight** – a shared memory supercomputer, operated by the Pittsburgh Supercomputing Center

## STORAGE RESOURCES AND NSF DATA PLANS

NCGAS and Indiana University will provide assistance to researchers seeking NSF funding or other funding that requires commitment of storage resources and data management plans. We offer:

- **Commitment of storage of your data in IU’s Scholarly Data Archive:** up to 50 TB and a minimum of three years, with two copies of data stored in Indianapolis and Bloomington, IN
- **Letters of commitment of storage resources** suitable for use as a part of your grant proposals
- **Assistance developing Data Management plans** as part of your proposals

## CASE STUDY:

### Melissa Pespeni’s Biodiversity Experiments

A specimen of *Onthophagus taurus*, one of the species of horned beetles Pespeni studies in her experiments on biodiversity.



Photo courtesy  
[http://commons.wikimedia.org/wiki/File:Onthophagus\\_taurus.JPG](http://commons.wikimedia.org/wiki/File:Onthophagus_taurus.JPG)

Melissa Pespeni is doing research about biodiversity in the animal kingdom, using horned dung beetles and sea urchins as model populations. NCGAS has provided services and expertise on the parts of the project relevant to the horned beetles, as described in this section.

Horned dung beetles have horns which may vary in size, position on the body, and sexual dimorphism between individuals. Rapidly evolving genes in *Onthophagus* may be involved in the diversity of expression of these traits. Pespeni is investigating the genomes of three species of *Onthophagus* for strong signals of selection.

NCGAS provided expert consulting in a number of ways to facilitate Pespeni’s experiments - recommending specific procedures and commands within UNIX, writing customized scripts to automate updating Pespeni’s data into the correct format for the selected analysis applications, troubleshooting issues, assisted with moving and verifying the accuracy of data, and adding and providing training for new users to the project.

## WANT TO KNOW MORE?

For more information about NCGAS, visit us on the web at <http://ncgas.org>

## Supported by...



PERVASIVE TECHNOLOGY  
INSTITUTE  
INDIANA UNIVERSITY

## Partners:

