



NATIONAL CENTER FOR GENOME ANALYSIS SUPPORT

INDIANA UNIVERSITY

a Pervasive Technology Institute (pti.iu.edu) Center

NCGAS is a national service center funded by the National Science Foundation's Advances in Biological Informatics (ABI) to provide scientists access to software and supercomputers for genomics research.

<http://ncgas.org>

NCGAS provides:

➤ **Dedicated access to memory rich supercomputers customized for genomics studies, including Mason and other XSEDE systems**

➤ **Distributions of hardened versions of popular applications**

- **Optimized Trinity now runs faster than other de novo assemblers**

➤ **Genome assembly software such as:**

- **de Bruijn graph methods: SOAPdeNovo, Velvet, ABySS**
- **consensus methods: Celera, Arachne 2**

Expanding to other areas as users are recruited: now moving into phylogenetics and metagenomics

We're especially interested in helping smaller institutions

Funded only in Nov. 2011, NCGAS is actively seeking users!



Mason Cluster:

- HP ProLiant DL580 G7
- 10GE interconnect
- rated at 3.383 TFLOPs (G-HPL benchmark)
- Quad socket nodes
- 8 core Xeon L7555 with 1.87 GHz base frequency
- 32 cores per node
- 512 GByte of memory per node!

Current participating institutions:

- Indiana University provides the Mason Cluster
- Texas Advanced Computing Center (TACC)
- San Diego Supercomputer Center (SDSC) provides DASH
- NCGAS will support software running at IU, TACC and SDSC, as well as other supercomputers available as part of XSEDE, with the goal to create a single allocation system that will transparently access all appropriate clusters
- NCGAS will further campus bridging integration



TEXAS ADVANCED COMPUTING CENTER
Powering Discoveries That Change The World



Thomas G. Doak (tdoak@indiana.edu), Richard LeDuc, Le-Shin Wu, Craig A. Stewart, Robert Henschel, William K. Barnett

<http://ncgas.org>

The First Six Months

In the first six months NCGAS has supported 21 separate NSF funded projects from West Virginia to Hawaii. These projects include data analysis support for techniques such as:

- Metagenomics
- Genome assembly and annotation
- RNA-sequencing for non-model organisms
- Proteomics

Organisms include *Paramecium* spp., *Stentor*, *Acanthamoeba*, *Oxytricha*, *petunia*, *flounder*, *Daphnia*, *copepods*, *salamanders*, *lizards*, *tomato*, *grasses*, and more. Metagenomic studies include the Ohio River aquatic ecosystem, and the human microbiome.

Partnered with the Trinity development community to optimize Trinity de novo assembler for high performance computing environments.

Partnering with CIPRES to facilitate interoperability of phylogenetic analysis.

Typical Projects

Genome Assembly and Annotation

Michael Lynch Lab (IU Bloomington, Department of Biology)

- Assemble and annotate 15 genomes in the *Paramecium aurelia* species complex and relevant out-groups, to study the evolutionary fates of duplicate genes. Used RNA-seq on each genome, to aid in genome annotation and to detect expression differences.
- Assemblies based on an overlap-layout-consensus method instead of a de Bruijn graph method. More memory intensive – requires performing pairwise alignments between all pairs of reads.
- Annotation of assemblies involves: GMAP, GSNAP, PASA, and Augustus. Requires load-in of millions of RNAseq and EST reads and maps them back to the genome.

Daphnia Population Genomics

Michael Lynch Lab (IU Bloomington Department of Biology)

- This project involves the whole genome shotgun sequences of over 20 diploid genomes with genomes sizes >200 Megabases each.
- With each genome sequenced to over 30 x coverage, the full project involves both the mapping of reads to a reference genome and the de novo assembly of each individual genome.
- The genome assembly of millions of small reads requires the excessive memory of Dash at SDSC and Mason at IU.

Collaborative Relationships

You need	We provide	Cost to you
Data analysis support <i>If your group has bioinformatics and HPC skills</i>	Free account <i>We provide a free account on the Mason server.</i>	Acknowledgement <i>Include NCGAS in your peer-reviewed publications</i>
Data analysis support <i>If your group lacks HPC skills</i>	Free account with access <i>We provide a free account on the Galaxy server with access to Mason.</i>	Acknowledgement <i>Include NCGAS in your peer-reviewed publications</i>
Support for a novel bioinformatics problem	Free consultation <i>We provide bioinformatics consultation and other services.</i>	Co-authorship <i>Intellectual contributions typically result in co-authorship</i>
Additional computational resources to propose a project to NSF	Letter of support <i>We provide documented support for use of NCGAS resources.</i>	Acknowledgement <i>Include NCGAS in your peer-reviewed publications</i>
Additional bioinformatics support to propose a project to NSF	Co-investigators <i>NCGAS staff can serve as co-investigators on a project.</i>	Sub-contracts <i>As needed within grants</i>

User Feedback

- "Access to the powerful servers of NCGAS has enabled us to get much better genome assemblies."
[Lukas Mueller](#)
[Boyce Thompson Institute for Plant Research](#)
- "NCGAS has provided high-end computing facilities to the *Petunia* Genome Project that were not otherwise available to our community. We are using NCGAS resources to carry out genome assemblies on a time frame significantly faster than would have been otherwise possible."
[Tom Sims](#)
[The Petunia Genome Project](#)
- "NCGAS is exactly the type of Bioinformatics support that small academic labs like ours need in order to make real progress on our genomics projects"
[Wendy Trzyna](#)
[Biological Sciences, Marshal University](#)

