



INDIANA UNIVERSITY

Today's Schedule

9am – 9:30am	Introduction
9:30-10:30	RNA-Seq Demo
10:30 – 11am	Workflows
11am – Noon	Lunch
1pm – 2pm	HPC overview
2pm – 5pm	Unstructured Galaxy Help

Introduction to Galaxy

Tom Doak

Le-Shin Wu

Carrie Ganote

National Center for Genome Analysis Support

August 12, 2015



INDIANA UNIVERSITY



Summary

- Who is NCGAS?
- Galaxy – what is it?
- Galaxy Anatomy



INDIANA UNIVERSITY

Who is NCGAS?

The National Center for Genome Analysis Support is based in Indiana but caters to a national audience with support from the NSF.

We provide computational resources and support for genomics, transcriptomics, and meta- projects.



INDIANA UNIVERSITY



PERVASIVE TECHNOLOGY
INSTITUTE

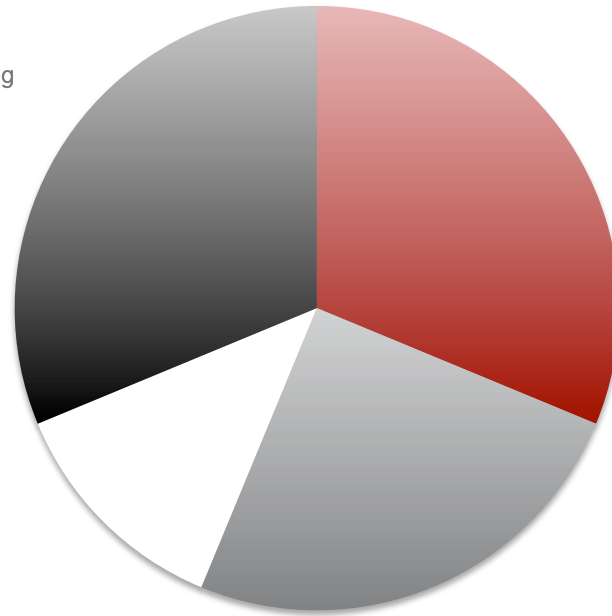
INDIANA UNIVERSITY

XSEDE

Extreme Science and Engineering
Discovery Environment



Open Science Grid



IU members

Le-Shin Wu

Carrie Ganote

Thomas Doak

Therese Miller

National Center for Genome Analysis Support: <http://ncgas.org>



What is Galaxy?

At its core, Galaxy is a web-based tool for analyzing data

- It provides a set of tools that one can apply to the data
- It provides a trail of analyses one has run
- It allows for sharing of data sets, methods, and workflows



What is Galaxy?

Galaxy was originally conceived as a tool to allow biologists to get work done without the need for training as an informatician

- Galaxy's purpose depends greatly on the tools made available to the user – these could be any command-line available software



INDIANA UNIVERSITY

There are many Galaxies in the Universe!

Galaxy Main – hosted at
Texas Advanced Computing Center



Galaxy-P – Proteomics-focused,
hosted at U of Minnesota

IU Galaxy – just another Galaxy
“instance” – each is autonomous
and specialized



NATIONAL CENTER FOR
GENOME ANALYSIS SUPPORT

INDIANA UNIVERSITY

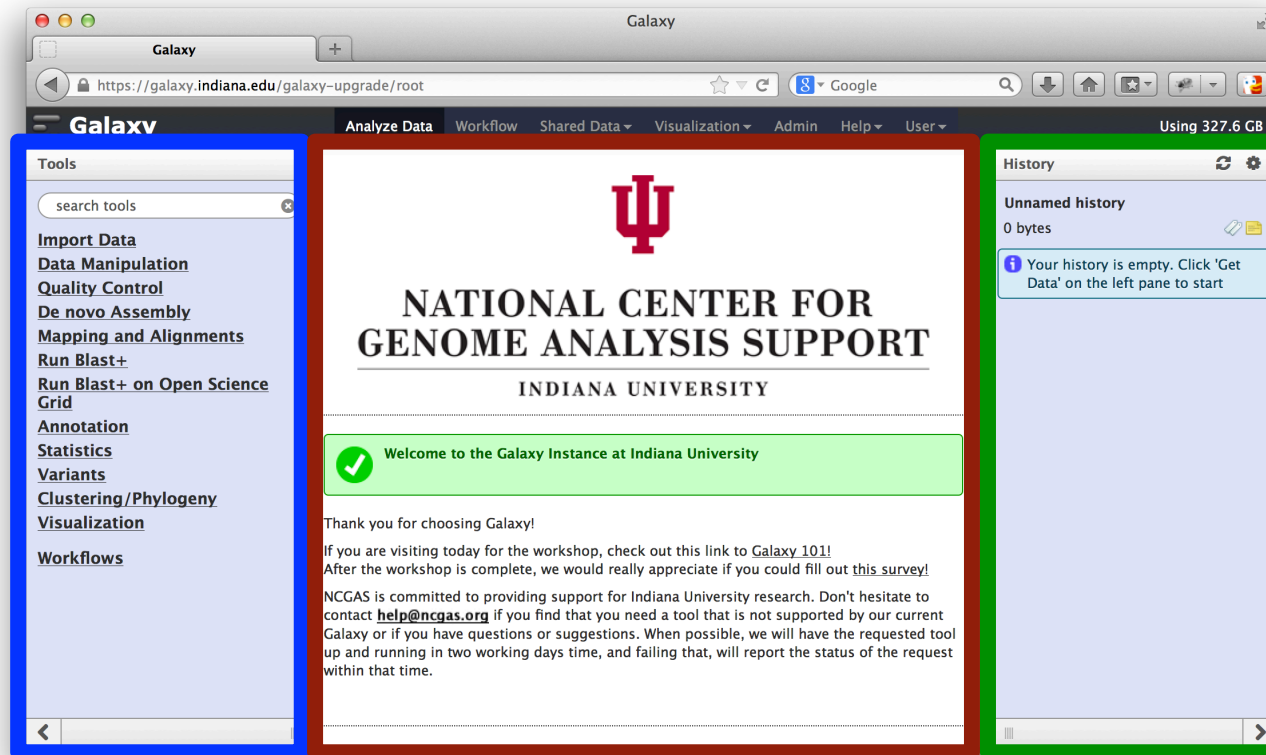
National Center for Genome Analysis Support: <http://ncgas.org>

Photo of M109 spiral galaxy courtesy of NOAO/AURA/NSF



INDIANA UNIVERSITY

Galaxy Anatomy and Physiology



Tool bar –
contains
the
available
steps to
apply to
data

History –
shows
steps
previously
taken to
manipulate
input data
sets

Focus pane – shows options,
parameters, and output for current item.

National Center for Genome Analysis Support: <http://ncgas.org>



Galaxy – the perfect technology?

There is no such thing as a perfect technology, but Galaxy might be good for you if you –

- Don't like looking up parameters/reading man pages
- Don't like managing job submissions/scripts
- Want an automatic trail of data analysis steps
- Like pastel colored web interfaces