State of the Galaxy 2014
GCC2012 | Chicago
First ever Galaxy Hackathon!

(See talk from Dannon Baker, Brad Chapman, John Chilton, Kyle Ellrott, et al)
1. State of the Community — GCC participation, code contributions, publications, Galaxy ToolShed et cetera

2. State of Galaxy main — move to TACC (more detail), statistics of Galaxy main usage

3. Some highlights from this last year — biostar, viz, dataset collections, toolshed, data managers, new UI: reference talks/lightning talks to come

4. Where we’re going — 10-ish priority areas

5. Acknowledgements
Community
(gcc, instances, code, toolshed, papers)
GCC participation

2010: 69
2011: 148
2012: 201
2013: 219
2014: 248
50+ public instances
12 Month Summary
Jun 30 2013 — Jun 30 2014

3738 Commits
Up +840 (28%) from previous 12 months

51 Contributors
Up +6 (13%) from previous 12 months
toolshed statistics

- 897 repositories
- 222 unique owners
- 176 Tool dependency package installation recipes
- 2,330 valid tools
- 3,420 valid versions of tools
- 54 exported Galaxy workflows
- 455 custom datatypes
- 62,021 total repository installations
All publications: cumulative by tag

Publications: Cumulative Tags per Year 2005-2013
vital signs of galaxy main
(infrastructure, jobs, users)
job numbers crash of 2013
22 May 2013: Initial proposal to move Galaxy main to TACC

29 June: Galaxy Team visits TACC to plan and hack

Mid August: Galaxy test running at TACC

October 7th: Galaxy main switched over to TACC completely

Continuing data migration in background…
CyberSTAR (NSF)
128 Cores
4 GB/core

BioSTAR (NSF)
128 Cores
8 GB/core

PSU Internal
10GB

Commodity Internet / Internet 2 / Lambda Rail
1 GB

Wartik 509

XSEDE Network

Data Supercell
Full Mirror

Pittsburgh Supercomputing Center

Galaxy Main Architecture: Extended
Galaxy Main at TACC

All physically co-located at the Texas Advanced Computing Center

Virtual machines

- Web Front-end 1
- Web Front-end 2
- Master Database
- Replicate Database

Corral
DDN Storage Appliance

Fileservers
926 TB of user data

Galaxy Cluster
256 Cores
16 GB/core
severe resource bottleneck
move to TACC
data is bigger = jobs are longer
highlights of the past year
(from Oslo to Baltimore)
Expansion to external resources
Viz framework
Dataset collections

Arbitrary # of Inputs (... paired).

Run applications in parallel (one per input).

Merged output for subsequent processing.
<actions>
  <action type="shell_command">git clone --recursive git://github.com/ekg/freebayes.git</action>
  <action type="shell_command">git checkout 9608597d12e127c847ae03aa03440ab63992fedf</action>
  <action type="shell_command">git submodule update --recursive</action>
  <action type="shell_command">make ll ( make clean & & sed -i.bak -e "s:LIBS = -lz -lm -L../vc/lib/tabixpp/"</action>
  <action type="move_directory_files">
    <source_directory>bin</source_directory>
    <destination_directory>$INSTALL_DIR/bin</destination_directory>
  </action>
</actions>
Figure S1. Schematic overview of reference data, location files, data tables, and tools.
A Galaxy Data Manager handles downloading, creating and installing each of the required facets necessary to provide built-in reference data through the use of a web-based GUI.
highlights of the future
(from Baltimore to ... )
This list contains items in order of priority as identified at March on the Baltimore team meeting.

They should include links to all cards that have been identified as parts of this effort.

- User Metrics + Analytics
  - Jeremy, James, Martin
  - RFC: Implement sophisticated user behavior analysis tool
  - Core: Fine-grained logging
  - User Metrics and Analytics

- UI for large scale analyses
  - John, Carl, Anton (user cases)
  - Dataset Collections - Models, API, Tools and Workflows
Complete niche-specific tools/workflows
New workflow scheduling engine
User Interface for true large-scale analysis
Visualization/Visual Analyses
ToolShed evolution
Cloud/Virtualization/Distributed Data/Federation
Scalable Training
Organizing Committee
Dave Clements, Johns Hopkins University
Mohammad Heydarian, Johns Hopkins University
Dan MacLean, The Sainsbury Laboratory
Karen Reddy, Johns Hopkins University

Scientific Committee
Jeremy Goecks, George Washington University
Jessica Kissinger, University of Georgia
Karen Reddy, Johns Hopkins University

Hackathon Committee
Dannon Baker, Johns Hopkins University
Brad Chapman, Harvard University
John Chilton, Penn State University
Kyle Ellrott, University of California Santa Cruz (UCSC)

Support
Stacey Hooker (GCC)
Teal Golden and Team (Housing)
Paula Davis (Hackathon)
thank you!

**Training Day Infrastructure**
- Dannon Baker
- Dave Bouvier

**Registration and Setup Volunteers**
- Eric Rasche
- Carl Eberhard
- Teresa Romeo Luperchio
- Xianrong (Jose) Wong
- Nick Stoler
- Martin Čech

**BoF organizers**
- Nikolay Aleksandrov Vazov
- Dave Clements
- Mo Heydarian
- Philip Blood
- John Chilton
- Nate Coraor
- Carrie Ganote
thank you!

Training Day Instructors
Jeremy Goecks (x2)
Aysam Guerler (x2)
Tom Bair
Jennifer Jackson
Nate Coraor
John Chilton
Simon Gladman
Andrew Lonie
Nikolay Vazov (x2)
Katerina Michalickova (x2)
Dannon Baker
Carl Eberhard
Greg Von Kuster

Training Day Instructors
JBjörn Grüning
Peter Cock
Saskia Hiltemann (x2)
Youri Hoogstrate (x2)
Hailiang (Leon) Mei (x2)
Jonas Paulsen
Tonje Lien Gulbrandsen
Morten Johansen
Karen Reddy
JJ Johnson
Dan Blankenberg
Ntino Krampis
Enis Afgan
Ravi Sanka
Brad Chapman