Agenda

Minimum Information About Galaxy to Get Going (MIAGGG)

Learning Galaxy with SNP/Variation Analysis

Galaxy Ecosystem (time allowing)

http://galaxyproject.org
What is Galaxy?

Data integration and analysis platform that emphasizes accessibility, reproducibility, and transparency

http://galaxyproject.org
What is Galaxy?

Keith Bradnam's definition:

"A web-based platform that provides a simplified interface to many popular bioinformatics tools."

From

"13 Questions You May Have About Galaxy"

Galaxy is available several ways ... 

http://galaxyproject.org
As a free for everyone service on the web: usegalaxy.org
A free for everyone web service:

http://usegalaxy.org

A free (for everyone) web server integrating a wealth of tools, compute resources, petabytes of reference data and permanent storage

However, *a centralized solution cannot support the different analysis needs of the entire world.*
Galaxy is available as Open Source Software

Galaxy is installed in locations around the world.

http://getgalaxy.org
Galaxy is available on the Cloud

http://aws.amazon.com/education
http://globus.org/
http://wiki.galaxyproject.org/Cloud
Galaxy on the Cloud: Galaxy CloudMan
http://usegalaxy.org/cloud

- Start with a **fully configured and populated** (tools and data) Galaxy instance.
- Allows you to scale up and down your compute assets as needed.
- Someone else manages the data center
Agenda

Minimum Information About Galaxy to Get Going (MIAGGG)

Learning Galaxy with SNP/Variation Analysis

Galaxy Ecosystem (time allowing)

http://galaxyproject.org
Quick Poll: Are you ...

1. A bioinformatics novice

2. A bioinformatics apprentice

3. A bioinformatics guru

Yes, those are your only choices.

http://galaxyproject.org
Demo Goals

Provide a basic introduction to using Galaxy for bioinformatic analysis using SNP calling as the driving example.

Demonstrate how Galaxy can help you explore and learn options, perform analysis, and then share, repeat, and reproduce your analyses.

If you happen to learn a little bit of bioinformatics and variant detection along the way, then that's a bonus.
SNP and Variation Analysis Live Demo

Demonstrate a variant analysis workflow
- get a public dataset
- check and maybe fix quality concerns
- map it
- identify variants
- determine effects

https://test.galaxyproject.org
Our data

- *Oryza sativa*
- Paired end DNA reads from an exome study
- Illumina HiSeq 2000
- From the **UC Davis Genome Center**
- Get our copy from EBI

- Using the full dataset, but it's relatively small
- No real science going on today!

http://www.ebi.ac.uk/ena/data/view/SRR1028565
SNP and Variation Analysis Live Demo

Lets do it.

https://test.galaxyproject.org
NGS Data Quality: Sequence bias at front of reads?

From a sequence specific bias that is 

caused by use of random hexamers in Illumina library preparation.

SNP and Variation Analysis: What we did

Get data from ENA
Examine quality with FastQC
Clean it up with Trimmomatic

Map it with Bowtie2
Removed unmapped and PCR dups with BAM Filter
Looked at mapped data with FastQC & IdxStats

Called variants with FreeBayes
Calculated effects with the Variant Effect Predictor @ EBI

https://test.galaxyproject.org
Agenda

Minimum Information About Galaxy to Get Going (MIAGGG)

Learning Galaxy with SNP/Variation Analysis

Galaxy Ecosystem (time allowing)

http://galaxyproject.org
2016 Galaxy Community Conference (GCC2016)

June 25-29, 2016
Bloomington, Indiana

galaxyproject.org/GCC2016
Galaxy Resources and Community

Mailing Lists (very active)
Unified Search
Issues Board
Events Calendar, News Feed
Community Wiki
GalaxyAdmins
Screencasts
Tool Shed
Public Installs
CiteULike group, Mendeley mirror
Annual Community Meeting

http://wiki.galaxyproject.org
Galaxy Community Resources: Galaxy Biostar

Tens of thousands of users leads to a lot of questions. Absolutely have to encourage community support.

Project traditionally used mailing list

Moved the user support list to Galaxy Biostar, an online forum, that uses the Biostar platform

Want help? Get answers.

https://biostar.usegalaxy.org/
Galaxy Community Resources: Mailing Lists
http://wiki.galaxyproject.org/MailingLists

**Galaxy-Dev**
Questions about developing for and deploying Galaxy
High volume (2336 posts in 2015, 1000+ members)

**Galaxy-Announce**
Project announcements, low volume, moderated
Low volume (36 posts in 2015, 6500+ members)

Also Galaxy-UK, -France, -Proteomics, -Training, ...
Unified Search: [http://galaxyproject.org/search](http://galaxyproject.org/search)

- Find
- Everything on ...
- Tools for ...
- Email about ...
- Source code for ...
- Published Histories, Pages, Workflows, about ...
- Related feature requests
- Papers using Galaxy for ...
- Documentation on ...
Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

- **Accessible**: Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible**: Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent**: Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

This is the Galaxy Community Wiki. It describes all things Galaxy.

**Use Galaxy**

Galaxy's public web server usegalaxy.org makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive user documentation applicable to any public or local Galaxy instance is available.

[usegalaxy.org](http://usegalaxy.org)

**Deploy Galaxy**

Galaxy is a free and open source project available to all. Local Galaxy servers can be set up by downloading the Galaxy application.

- Admin
- Cloud

[getgalaxy.org](http://getgalaxy.org)

**Contribute**

Galaxy has a large and active user community and many ways to get involved.

- **Community**
- **Users**: Share your histories, workflows, visualizations, data libraries, and Galaxy Pages, enabling others to use and learn from them.
**Events**

**Galaxy Event Horizon**

Events with Galaxy-related content are listed here.

Also see the Galaxy Events Google Calendar for a listing of events and deadlines that are of interest to the Galaxy Community. This is also available as an RSS feed.

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar, please send it to outreach@galaxycloud.org.

For events prior to this year, see the Events Archive.

**Upcoming Events**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic/Event</th>
<th>Venue/Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 12</td>
<td><strong>Introduction to Galaxy Workshop</strong></td>
<td>Virginia State University, Petersburg, Virginia</td>
</tr>
<tr>
<td>December 16-19</td>
<td><strong>RNA-Seq and ChIP-Seq Analysis with Galaxy</strong></td>
<td>UC Davis, California, United States</td>
</tr>
<tr>
<td>January 10-14</td>
<td><strong>Galaxy for SNP and Variant Data Analysis</strong></td>
<td>Plant and Animal Genome XXIII (PAG2014), States</td>
</tr>
<tr>
<td>January 19-20</td>
<td><strong>NGS pipelines with Galaxy</strong></td>
<td>e-Infrastructures for Massively Parallel Sequencing, Sweden</td>
</tr>
<tr>
<td>February 9-13</td>
<td><strong>Analyse bioinformatic de séquences sous Galaxy</strong></td>
<td>Montpellier, France</td>
</tr>
<tr>
<td>February 16-18</td>
<td><strong>Accessible and Reproducible Large-Scale Analysis with Galaxy</strong></td>
<td>Genome and Transcriptome Analysis, Conference, San Francisco, California</td>
</tr>
<tr>
<td></td>
<td><strong>Large-Scale NGS data Analysis on Amazon Web Services Using Globus Genomic</strong></td>
<td>Genomics &amp; Sequencing Data Integration, of Molecular Medicine Tri-Conference, San Francisco, California</td>
</tr>
</tbody>
</table>

**News**

**News Items**

**Opening at McMaster University**

The McArthur Lab in the McMaster University Department of Biochemistry & Biomedical Sciences is seeking a Systems Administrator / Information Technologist to help establish a new bioinformatics laboratory at McMaster, plus develop the next generation of the Comprehensive Antibiotic Resistance Database (CARD).

From the job announcement on EvolDir:

The candidate will configure BLADE and other hardware for general bioinformatics analysis, development of a GIT version control system, construction of an in house Galaxy server (usegalaxy.org), and development of a new interface, stand-alone tools, APIs, and algorithms for the CARD (based on Chado).

See the full announcement for details.

**December 2014 Galaxy Newsletter**

As always there’s a lot going on in the Galaxy this month. “Like what?” you say. Well, read the dang December Galaxy Newsletter we say! Highlights include:

- Galaxy Day! In Paris! This Wednesday!
- Near Richmond, Virginia? There’s a Galaxy Workshop at Virginia State U on December 12.
- GCC2015 needs sponsors!
- Other upcoming events on two continents
- **96 new papers**, including 6 highlighted papers, referencing, using, extending, and implementing Galaxy.
- Job openings at 7+ organizations
- A new mailing list: Galaxy-Training
- 15 new ToolShed repositories from 10 contributors
- And, 10 other juicy (well maybe not juicy, but certainly not crunchy) bits of news

Dave Clements and the crisp Galaxy Team

**Bioinformatics, Freiburg**

Max Planck Institute of Immunobiology and Epigenetics in Freiburg, Germany has an opening for a Bioinformatician for an initial period of two years. The successful candidate will work at the interface between an in-house deep sequencing facility (HiSeq-2500) and the various research groups at the institute. Main responsibilities include primary analysis of deep sequencing data and quality control.
Community can create, vote and comment on issues

We also support community organized efforts and events.
Galaxy Resources & Community: Videos

“How to” screencasts on using and deploying Galaxy

Talks from previous meetings.

http://vimeo.com/galaxyproject
Galaxy Resources & Community: CiteULike Group

Now almost 3000 papers

Galaxy Training Network launched in October.

bit.ly/gxygtn
Galaxy Project: Further reading & Resources

http://galaxyproject.org
http://usegalaxy.org
http://getgalaxy.org
http://wiki.galaxyproject.org/Cloud
Further adventures in Galaxy

Galaxy Community Update

Wednesday 11:25, in Golden West
Covering recent enhancements and activity in the Galaxy community.

Part of the GMOD workshop that starts @ 10:30
The Galaxy Team

Enis Afgan                   Dannon Baker       Dan Blankenberg       Dave Bouvier       Marten Cech
John Chilton

Dave Clements                Nate Coraor         Carl Eberhard         Jeremy Goecks      Sam Guerler

Jen Jackson                  Ross Lazarus        Anton Nekrutenko      Nick Stoler         James Taylor       Nitesh Turaga

http://wiki.galaxyproject.org/GalaxyTeam
Acknowledgements

You
Anthony Bolger
Nate Coraor
PAG
NIH
Johns Hopkins University
Penn State University