

NGS Data Analysis and Galaxy

University of Pretoria
Pretoria, South Africa
14-18 October 2013

Dave Clements, Emory University

<http://galaxyproject.org/>

Fourie Joubert, Burger van Jaarsveld
Bioinformatics & Computational Biology Unit
University of Pretoria

<http://science.up.ac.za/>



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



This Week

Monday	Welcome, Project Intro, Basic Galaxy Usage NGS QualityControl
Tuesday	RNA-Seq - Mapping and Transcript Prediction RNA-Seq: Differential expression and Alternative Pipelines; SNP & Variant Analysis
Wednesday	SNP & Variant Analysis Chip-Seq Analysis
Thursday	Genome Assembly Install your own Galaxy on Amazon Cloud
Friday	Customizing Galaxy, Galaxy Tool Shed, and Wrapping Tools for Galaxy

Acknowledgements

Pretoria

Jessika Samuels

Fourie Joubert

John Becker

Burger van Jaarsveld

John Ambler

Cape Town

Nikola Mulder

Gerrit Botha

Cashifa Kerriem

Atlanta

Dannon Baker

Tutorials & Datasets

RNA-Seq

Monica Britton,

Nikhil Joshi, Joe Fass

ChIP-Seq

Shannan J. Ho Sui,

Oliver Hoffman

Assembly

David Edwards,

Kathryn Holt

Acknowledgements



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Friday Agenda

- 8:30 **Welcome and Questions**
- 8:45 **Intro to the Linux command line**
- 9:45 **Galaxy Revealed! Part I**
- 10:30 Break
- 11:00 Galaxy Revealed! Part II
- 12:30 Lunch
- 13:30 Galaxy Tool Wrapping Part I
- 15:00 Break
- 15:30 Galaxy Tool Wrapping Part II
- 16:30 Done

Create a new Cloud Instance

<http://bit.ly/gxyup13>

Save your PEM file!

Create an account on the server
and make it an admin through the
CloudMan interface.

SSH Magic

```
$ ssh -i whatever.pem \  
    ubuntu@ec2-xxx-xxx-xxx-xxx.region.compute.amazonaws.com  
The authenticity of host 'ec2-xxx-xxx-xxx-  
xxx.region.compute.amazonaws.com (xxx-xxx-xxx-xxx)' can't be  
established.  
RSA key fingerprint is  
xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx:xx.  
Are you sure you want to continue connecting (yes/no)? yes  
Warning: Permanently added 'ec2-xxx-xxx-xxx-  
xxx.region.compute.amazonaws.com' (RSA) to the list of known  
hosts.  
Welcome to Ubuntu 12.04.3 LTS (GNU/Linux 3.2.0-53-virtual  
x86_64)  
...  
ubuntu@ip-xxx-xxx-xxx-xxx:~$
```

And now you are in the command line shell of your
Galaxy instance

A word on users

```
$ ssh -i whatever.pem \  
  ubuntu@ec2-xxx-xxx-xxx-xxx.region.compute.amazonaws.com  
...  
ubuntu@ip-xxx-xxx-xxx-xxx:~$
```

```
ubuntu$  
$
```

“ubuntu” at front of prompt means you are now logged in as the ubuntu user. We’ll spend most of our time as one of these:

ubuntu:

Has the ability to *invoke superpowers and to become other users*

galaxy:

Owens the Galaxy installation and files.

A word on *sudo*

```
ubuntu$ a command ubuntu does not have permission for
Permission denied.
ubuntu$ sudo same command
ubuntu$

ubuntu$ sudo su -l galaxy
galaxy$ sudo anything
Permission denied.
galaxy$ exit
ubuntu$
```

The `sudo` command allows certain users to do things normal users can't do, including becoming other users.

ubuntu can run `sudo`.

galaxy cannot.

Where things are

```
$ cd /mnt
$ ls
cm galaxy galaxyData galaxyIndices lost+found transient_nfs
$
```

cm

CloudMan details

galaxy

The running Galaxy instance. We'll spend most of our time here

galaxyData

Empty until something happens

galaxyIndices

All the reference data that comes packaged with Galaxy on the cloud

The Universe!

```
$ sudo su -l galaxy
$ cd /mnt/galaxy/galaxy-app
$ less universe_wsgi.ini
$ pico universe_wsgi.ini
$
```

universe_wsgi.ini

Galaxy instance configuration details

Add a brand to your server.

In the [app:main] section, add

```
brand = My Cool Brand
```

Restart Galaxy through CloudMan admin page and then refresh your main Galaxy page.

Friday Agenda

- 8:30 **Welcome and Questions**
- 8:45 **Intro to the Linux command line**
- 9:45 **Galaxy Revealed! Part I**
- 10:30 **Break**
- 11:00 **Galaxy Revealed! Part II**
- 12:30 **Lunch**
- 13:30 **Galaxy Tool Wrapping Part I**
- 15:00 **Break**
- 15:30 **Galaxy Tool Wrapping Part II**
- 16:30 **Done**

Friday Agenda

- 8:30 **Welcome and Questions**
- 8:45 **Intro to the Linux command line**
- 9:45 **Galaxy Revealed! Part I**
- 10:30 **Break**
- 11:00 **Galaxy Revealed! Part II**
- 12:30 **Lunch**
- 13:30 **Galaxy Tool Wrapping Part I**
- 15:00 **Break**
- 15:30 **Galaxy Tool Wrapping Part II**
- 16:30 **Done**

Add a reference genome: 1st step

Define the DB KEY:

```
$ cd /mnt/galaxy/galaxy-app/tool-data/shared/ucsc/  
$ pico builds.txt  
$ pico universe_wsgi.ini  
$
```

Add

```
milliv      Milli Vanilli (milliv)
```

as the last line. Note, that is a **TAB, not spaces**, between columns 1 and 2.

Save the file.

Add a reference genome: Data Managers

<http://bit.ly/gxyDataManagers>

<http://bit.ly/gccDataManagers>

Cut and paste and follow directions on HowTo page.

Use defaults

```
galaxy$ cd /mnt/galaxy/galaxy-app/tools
$ mkdir data_manager
$ cd data_manager
$ pico fetch_genome_all_fasta.xml
$ pico data_manager_fetch_genome_all_fasta.py
$ cd ../..
$ pico data_manager_conf.xml
$ pico universe_wsgi.ini
```

Restart Galaxy

Add a reference genome: Data Managers

On your **Cloud[123].galaxyproject.org**

Shared Data → Published Histories → Milli v 101

Click on [m.vannielli.sequence.fasta](#)

Hover over the diskette icon.

Copy the URL

On your **own cloud instance**

Analyze Data → Get Data

→ Upload File

Add a reference genome: Data Managers

On your cloud instance, using the Galaxy GUI

Admin → Data → Manage Local Data

Add a reference genome: Data Managers

On your cloud instance, using the Galaxy GUI

Admin → Data → Manage Local Data

Where is it? Let's look at the log.

```
galaxy$ cd /mnt/galaxy/galaxy-app/  
$ less main.log  
  
galaxy.tools.data_manager.manager ERROR 2013-10-18  
05:53:55,807 There was an error parsing your Data Manager  
config file "./data_manager_conf.xml": mismatched tag: line  
19, column 6  
  
$ nl data_manager_conf.xml  
$ nano data_manager_conf.xml
```

Restart Galaxy

Add a reference genome: Data Managers

On your cloud instance, using the Galaxy GUI

Admin → Data → Manage Local Data

→ Access Data Managers

→ Reference Genomes

Source → History or URL

Add a reference genome: Data Managers

Cool!

but...

NGS: Mapping → Bowtie

Ain't no Milli V.

Add a reference genome: Data Managers

Bowtie has its own indexes for genomes.
Lots of tools require preprocessing steps.

Need to tell Bowtie about Milli V.

Admin → Search and Browse Tool Sheds
Select Galaxy Test Tool Shed

Search for bowtie

Select data_manager_bowtie_index_builder

Click Install to Galaxy

Add a reference genome: Data Managers

Repository installation is not possible due to an invalid Galaxy URL: None. You may need to enable cookies in your browser.

Well, how did we do it back in 2012?

```
galaxy$ cd /mnt/galaxy/galaxy-app/tool-data/milliV/seq
$ ls
$ ../../../../../../tools/bowtie/default/bin/bowtie-build -f milliV1.fa milliV1
$ mkdir ../bowtie
$ mv milliV1.*ebwt ../bowtie/
$ pico /mnt/galaxyIndices/galaxy/tool-data/bowtie_indices.loc

milliV1      milliV1      milliV1: Milli Vanilli      /mnt/galaxy/galaxy-app/tool-
data/milliV/bowtie/milliV1
```

Make sure those are TABs, not spaces.

Add a Tool: htseq-counts

We used htseq-counts to create matrices for DESeq

Admin → Search and Browse Tool Sheds
Select Galaxy Main Tool Shed

Search for htseq

Click Install to Galaxy

One more tool: DESeq

DESeq was the CuffLinks alternative we used this week

Admin → Search and Browse Tool Sheds
Select Galaxy Main Tool Shed

Search for DESeq

Click Install to Galaxy

Thanks



Dave Clements

**Galaxy Project
Emory University**

clements@galaxyproject.org