A Galaxy of learning:

Bioinformatics tutorials based on Galaxy

Simon Gladman
Background and the Genomics Virtual Lab
We come from a land down-under
The Genomics Virtual Lab

Nationally funded and distributed platform for genomic analyses

Genomics Virtual Lab

- Tools
- Workflows
- Visualisation
- Tutorials

Australian Research Cloud Infrastructure

Galaxy

UCSC Browser

Bio Linux
The Genomics Virtual Lab

- Sets of scripts and machine images for building GVL instances
  - Including Galaxy and CloudBioLinux
  - Reference datasets
- Can run on:
  - Stand alone systems
  - Cloud via Cloudman
- Public Galaxy service for training
  - galaxy-tut.genome.edu.au
- A suite of tutorials and protocols for common genomics tasks
Training Using Galaxy
Galaxy Tutorials

Current:
- *de novo* Assembly
- RNA-Seq DGE (Basic)
- Variant detection (Basic)
- Variant detection (Advanced)

Forthcoming:
- Variant detection (Somatic)
- RNA-Seq DGE (Advanced)
- Transcriptome assembly
- Metagenomics
- CHiP-Seq
- Bacterial Genome Annotation
Galaxy Tutorials

- Layered Google documents
- Use real datasets
- [galaxy-tut.genome.edu.au](http://galaxy-tut.genome.edu.au)
- Histories at each step
- Galaxy server image forthcoming

Step by Step instructions

Background

Tutorial Document

Screenshots
Protocols

- Less prescriptive than Tutorials
- Presents concepts of an analysis
- Suggests tools and alternates for each step:
  - Galaxy
  - Command-line
- Includes substantial background
Workshops

- One day workshop per topic
- Seminar/lectures in Morning
- Afternoon
  - Hands on with Galaxy
  - Tutorial walkthrough
  - Q&A on analysis with Galaxy
- Feedback has been extremely positive
- Community of Galaxy users is growing
Tutorial Development
Tutorial development

**Analysis design & tool selection**

- Domain experts:
  - Design of analysis
  - Tool selection
  - Parameter selection
  - Sample data selection

- Bioinformatician/Galaxy dev:
  - Wrapper design
  - Implementation
  - Installation on GVL image
Tutorial development

- **Domain experts:**
  - Run analysis on data
  - Snapshot histories at each step
  - Tweak tool interfaces

- **Team:**
  - Write tutorial document
  - Collate and link background
  - Review and publish on web
Development issues

- Tool wrapping
  - Level of parameter exposure in wrapped tools
  - Test suite versus tool wrapper capabilities

- Tool dependencies and binary installations: "Shrink Wrapping"
  - Dependency handling is confusing
  - Would be good to see some more complex examples
  - Not suitable for all applications - e.g. Qiime
  - Low level dependencies - e.g. Kernel headers, BioPerl etc
  - Where to draw the line between machine/toolshed
  - Galaxy-dev mailing list is really helpful!
Updates/future work

● Keeping the tutorials up to date
  ○ Adding new tools as they become standard
  ○ Updating documentation to match

● Future/continuing work
  ○ New toolshed mechanisms and the toolshed police
  ○ Adapt to changing Galaxy landscape
  ○ Automated testing for all tools
  ○ "Shrink wrap" tool dependencies into tool repositories
Conclusions
Has anyone learned anything?

- We've learned how to set up Galaxy and wrap tools
- We are working on "Shrink Wrapping"

Most importantly:
- People who have used the tutorials / protocols / workshops have learned some useful Bioinformatics analysis techniques..
- And their level of fear regarding these analyses has been reduced!
- Lots of positive feedback.
The team

- **VLSCI**
  - Andrew Lonie
  - Clare Sloggett*
  - Enis Afgan*
  - Nuwan Goonasekera
  - Simon Gladman*
  - Mahtab Mirmomeni
  - Franco Caramia

- **University of Queensland**
  - Ron Horst*
  - Igor Makunin

- **Garvan Institute**
  - Warren Kaplan
  - Kevin Ying*
  - Derrick Lin

- **CSIRO**
  - Phillippe Moncuquet*

- **QFAB**
  - Mark Crowe
  - Pierre-Alain Chaumeil*
  - Xinyi Chua
  - Anne Kunert
Thank you.