

Building a user community for Galaxy

Annette McGrath | Bioinformatics Core Leader
25 March 2014

CSIRO COMPUTATIONAL INFORMATICS
www.csiro.au



The CSIRO Galaxy Pilot

- What was the motivation?
 - To see if there was a demand for a Galaxy service within CSIRO
- The Galaxy Pilot project began in late 2011
- Announced in September 2012
- A successful collaboration between the Bioinformatics Core and IT
- Following very positive feedback from users, we are moving to a full production service

The impact of training and support on the uptake of the CSIRO Galaxy pilot

The Galaxy pilot team

- A fluid group of people from both IT and bioinformatics
- The team
 - Annette McGrath
 - Steve McMahon
 - Philippe Moncuquet
 - Sean Li
 - Ondrej Hlinka
 - Contributions by Josh Bowden, Michael James, Brian Davis
 - Contributions by Darren Cullerne, Sean McWilliam, Lauren Bragg, Neil Saunders
 - Now joined by Tim Ho and Joel Ludbey-Bruhvel

Galaxy / CSIRO

galaxy.bioinformatics.csiro.au

Galaxy / CSIRO Analyze Data Workflow Shared Data Visualization Admin Help User Using 2%

Tools

- send data
- ENCODE Tools
- Text Manipulation
- Unix Tools
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- FASTA manipulation
- CSIRO TOOLS
- BioKanga 2.9.9
- blue
- Acacia error-correction
- NGS utilities
- Linkage-cw
- STATISTICS
- Stats
- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Motif Tools
- ALIGNMENT
- NCBI BLAST+
- Multiple Alignments
- ENVIRONMENTAL GENOMICS
- Metagenomic analyses
- Qiime 1.6.0
- NGS TOOLBOX
- NGS: Picard (beta)
- NGS: QC and manipulation

History

Unnamed history

0 bytes

Your history is empty. Click 'Get Data' on the left pane to start

This instance of Galaxy is supported by the CSIRO Bioinformatics Core and IM&T

To get started please visit <https://wiki.csiro.au/display/BioinformaticsCore/Getting+started>

You can find useful workflows/datasets/tutorials in the 'Shared data' section, you can also find helpful guides in the 'Help' section

For assistance, tools and genomes request, please submit a ticket by emailing: galaxyhelp@hpsc.csiro.au

For the best experience possible of Galaxy we recommend to use Chrome or Firefox

The CSIRO Galaxy pilot now

- 462 tools installed, mostly NGS, some custom
- 227 registered users
- Monthly usage fluctuates but averages about 1750-2000 jobs a month

Who are the users

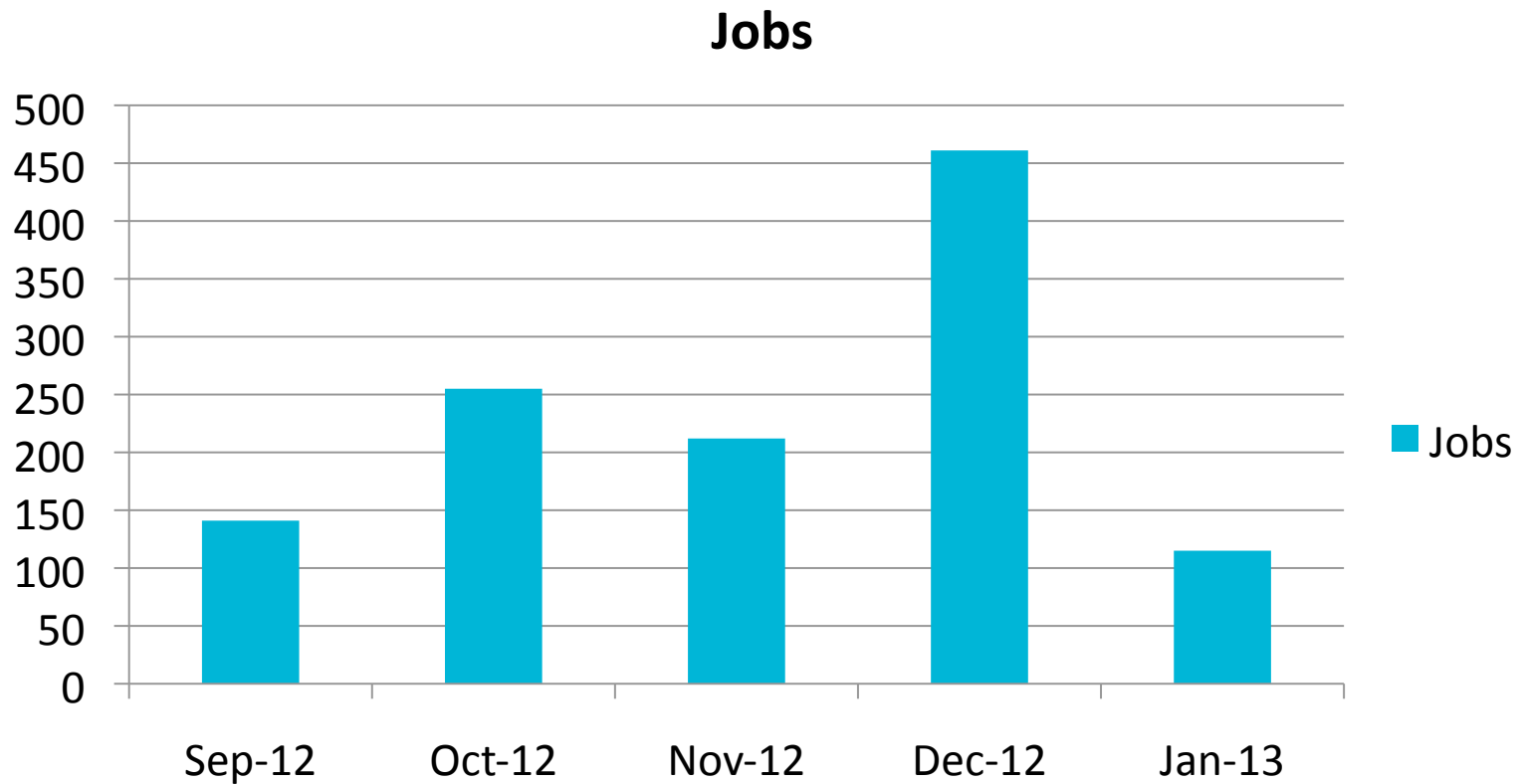
- CSIRO currently has ~6000 staff in 52 locations in 11 divisions.
- Users from
 - CSIRO Plant Industry
 - CSIRO Animal Food and Health Sciences
 - CSIRO Marine and Atmospheric Research
 - CSIRO Land & Water
 - CSIRO Ecosystem Sciences
 - CSIRO Material Science and Engineering
 - CSIRO Computational Informatics

The ecosystem of bioinformatics solutions available



Announcement and uptake of the new service

- Announced in the weekly CSIRO newsletter to all staff – Sept 2012



Early training activities

- Galaxy basics on our wiki and communicated to users
 - The interface
 - How to get data into Galaxy
 - How to do a simple analysis – BLAST
 - How to do QC on your NGS data
 - How to make a workflow
 - How to share your data/history
- We ran some online demos –MeetingPlace
- BPA/CSIRO/EBI NGS hands-on training workshop

Contents

- Home
 - How to
 - Planning
- Projects
 - Galaxy Pilot
 - Galaxy developer information
 - Galaxy user information
 - Getting started
 - Interface
 - Uploading data into Galaxy

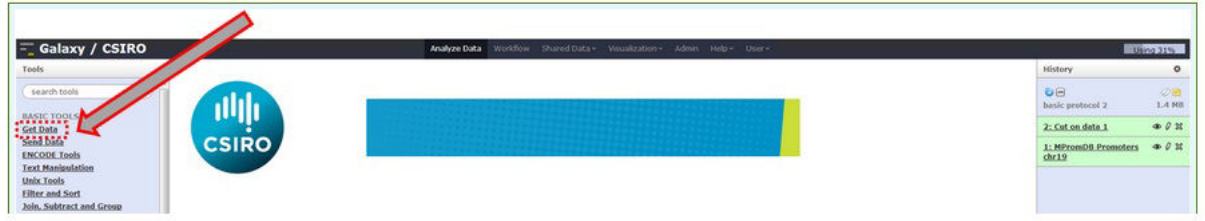
Uploading data into Galaxy

updated Mar 13, 2013 by Hlinka, Ondrej (CSIRO IM&T, St. Lucia)

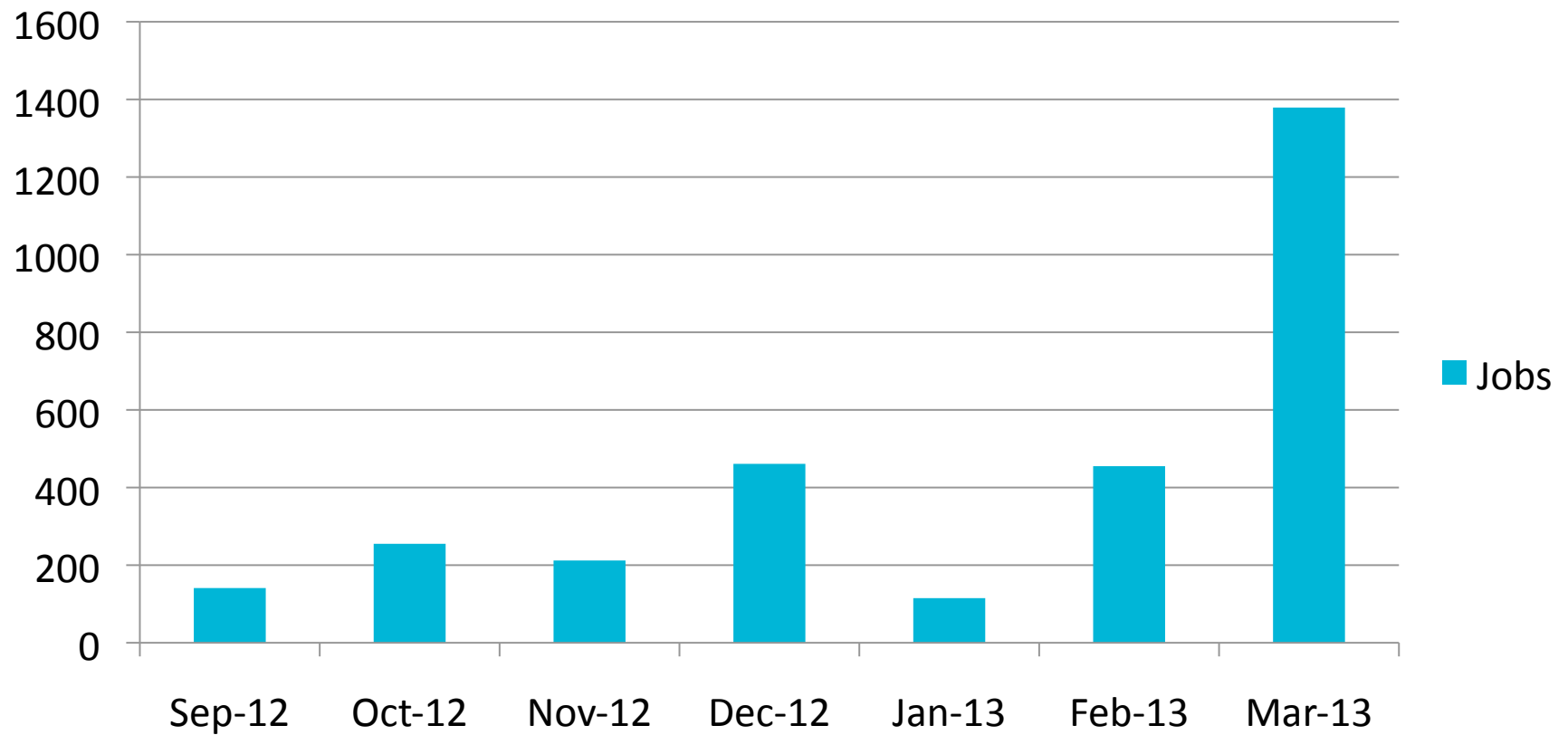
How do I get my data into galaxy

1- Upload from computer (the classic method)

Click on "Get Data"



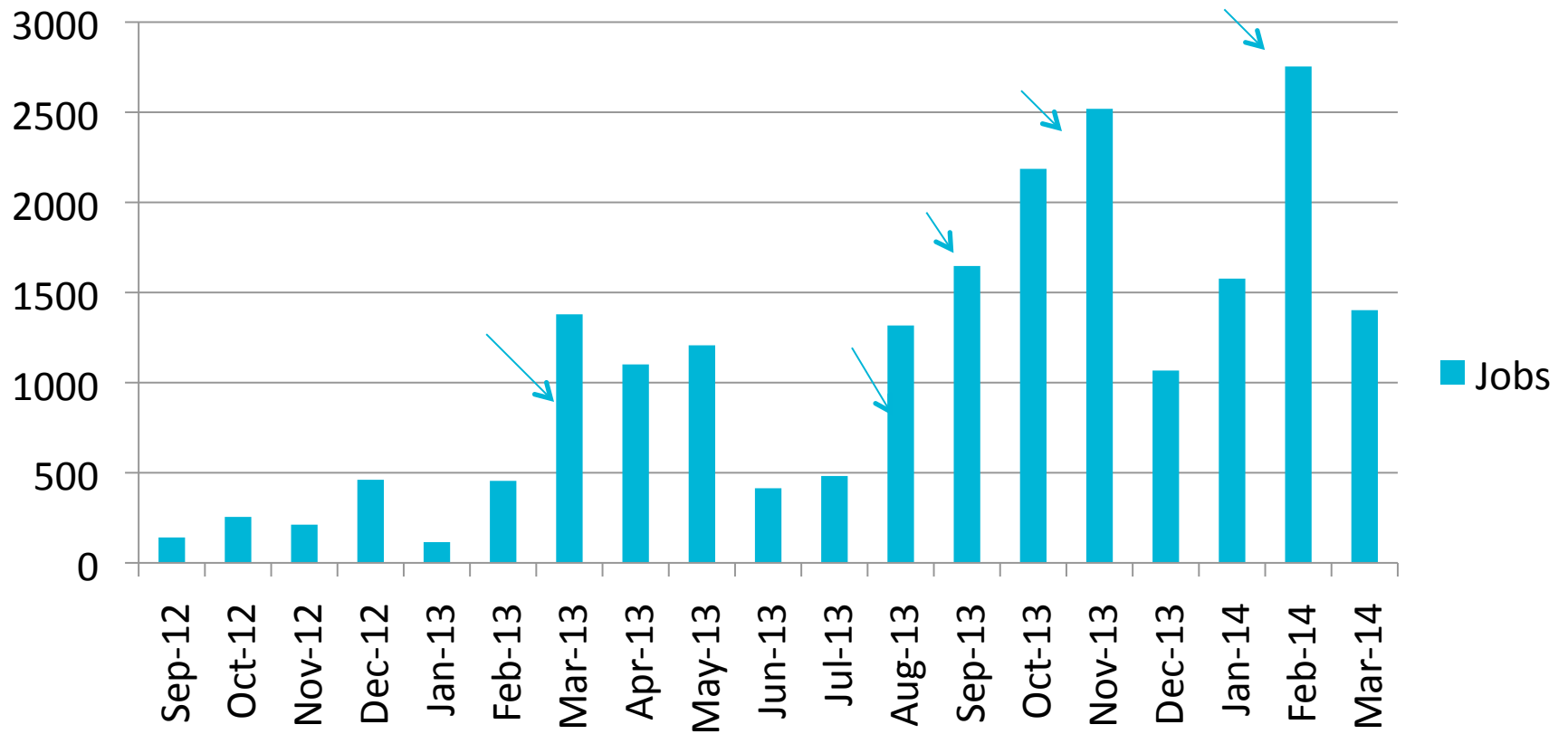
Jobs



Other training activities

- Intro to RNASeq – command line
- RNASeq using Galaxy
- 2 day Using CSIRO's Galaxy for RNASeq analysis
 - Mixture of theory and plenty of practice
 - Training material
 - Galaxy 101
 - Material on usegalaxy.org
 - GVL
- More NGS workshops
- BPA/CSIRO/EBI metagenomics workshops in Feb 2014

Jobs per month



What type of support queries

- A survey of this year's tickets :
 - Advanced output options for blastn
 - Requests for metagenomics datasets and tools
 - Removal of MIDs and primers from sequence reads (non-standard)
 - Creation of a workflow for a user
 - Query regarding sam/bam output of velvet
 - Subsetting data from a reference dataset
 - Reporting a bug which turned out to be due to IE usage
 - Creating custom subsets of nr
 - Gff to gtf issues
 - Implementing tools for clustering in Galaxy
 - Advice re use of tools for DE
 - Advice on creating custom blast dbs

How is support handled

- Access to a specific queue on the hpsc RT system
- Bioinformaticians see the queries as well as a group of IT staff and sys admins
- Queries tend to fall into the categories of
 - I got a funny result, why?
 - How can I do x in galaxy using tool y?
 - I want to do X simple task, can you create a tool or a workflow for me?
 - I'm having problems with file formats
 - I saw this great new tool! Can we put it in galaxy and compare it to xx ..
 - Can you increase my quota
 - I just want all of the archeael 16S genes
 - From bioinformaticians – we wrote this tool ages ago in X and now it may be best access via Galaxy

Domain knowledge is essential. Resolution often requires input from IT & bioinformaticians

Impact of support

- Users know that if they can try Galaxy and if they get stuck that they will get help – there is a safety net
- Help can take many forms
- Engenders a great deal of goodwill towards the service

Future directions

- You will hear more from Steve about our planned production service
- We will continue to add more tools – stat gen and metagenomics are promising areas
- We will continue to support users and are developing more training course
- Integration with the Bioinformatics Reference Data project
- Investigating the use of gpu-based bioinformatics tools and possible integration with Galaxy

Conclusions

- These are the lessons learned from our Galaxy pilot
- An interdisciplinary implementation team worked very well
- Build it and they will come?
 - With some encouragement!
 - Training was essential to the uptake of the service
 - Raising awareness of the Galaxy can do
 - Responsive support has also played a major role in gaining trust and acceptance for the service.
 - Domain knowledge is essential
 - Acting on a support request may not always be the best answer

Now, new starters sign up to Galaxy at the same time they request a CLCBio or Geneious licence – success!

Thank you

**CSIRO COMPUTATIONAL
INFORMATICS/Bioinformatics Core**

Annette McGrath
Bioinformatics Core Leader

t +61 2 6216 7251

e annette.mcgrath@csiro.au

ADD BUSINESS UNIT/FLAGSHIP NAME
www.csiro.au

