GalaxyP: A Galaxy Proteomics Community

Ira Cooke
What is GalaxyP?

**Biochemistry, Molecular Biology & Biophysics**
- TIM GRIFFIN
- Candace Guerrero
- Praveen Kumar

**Biologists / collaborators**
- Katie Vermillion
- Kyle Anderson
- Matt Andrews
- Brett Noel, PhD
- Tzu-Yi Yang, PhD
- Laurie Parker
- Brian Sandri
- Alexa Pragman
- Chris Wendt
- Somiah Afuni
- Kristin Boylan
- Marnie Peterson
- Amy Skubitz
- Kevin Viken
- Maneesh Bhargava

**Minnesota Supercomputing Institute**
- Michael Milligan
- Tom McGowan
- Getiria Onsongo
- JAMES JOHNSON

**Galaxy Team**
- John Chilton
- Mo Heydarian
- Karen Reddy

**Center for Mass Spectrometry & Proteomics**
- Ebbing de Jong
- LeeAnn Higgins
- Todd Markowski

**Funding**
- National Cancer Institute
- National Science Foundation

**Harald Barsnes & Marc Vaudel**
- University of Bergen, Bergen, Norway

**Bjoern Gruening**
- University of Freiburg, Freiburg, Germany

**Ira Cooke**
- La Trobe University, Melbourne, Australia
Part 1: Proteomics before Galaxy-P
In the bad old days ... everything was:

Monolithic
Proprietary
Inextensible
Windows only
A few pioneering open source tools appeared ...

Trans Proteomic Pipeline

Proteowizard

X!Tandem

OpenMS

along with open data formats
Standardization efforts made more formats ...
Doing real work with these tools was very messy
Part 2: Galaxy to the Rescue
We built lots of tools

Added Proteomics Category

90 Tool Suites

Hundreds of individual Tools
We made tools available on public servers

https://usegalaxyp.org
We made graphical interfaces
We added lots of new Datatypes to Galaxy
We made some “crazy” but useful workflows
We also got some real science done.
Did we really improve things?
YES
A complex workflow in galaxy is better than a complicated shell script

Workflows have many hidden benefits

- Expose bad design choices
- Expose incompatibilities between tools
- Expose bugs and unreliable tools
- Expose the inner workings of the method

+ Sharable, reproducible etc etc
Case Study: PeptideShaker + SearchGUI
SearchGUI and PeptideShaker

Desktop GUI interface ideal for small datasets

CLI on a cluster required for realistic sized datasets

CLI gave us hope for a Galaxy Tool

Marc Vaudel

Harald Barsnes
Hundreds of small files
Related to each other by full paths
GalaxyP improved these tools

*Bugs in multi-user environments exposed

*Improved packaging of results files
Part 3: Github
Github as a centre of development

GalaxyP
Galaxy Tool Shed repositories maintained and developed by the GalaxyP community

🔗 https://github.com/galaxyproteomics/

13 Contributors

82 Pull Requests

Hackathon (Dec 2016) focussed on metaproteomics
A snapshot of activity (January 2017)

January 2, 2017 – February 2, 2017

Overview

13 Active Pull Requests

12 Merged Pull Requests

1 Proposed Pull Request

1 Closed Issue

3 New Issues

Excluding merges, 7 authors have pushed 51 commits to master and 51 commits to all branches. On master, 54 files have changed and there have been 1,923 additions and 1,031 deletions.
Building Trust in the Tools

Add more commits by pushing to the `msp` branch on `galaxyproteomics/tools-galaxyp`.

- **All checks have passed**
  2 successful checks

- **This branch has no conflicts with the base branch**
  Merging can be performed automatically.

- **Merge pull request**
  You can also open this in GitHub Desktop or view command line instructions.
An active, diverse and vocal community helps!

Build better tools

Enable new science