Using Galaxy for Proteogenomics

Ira Cooke
What can Proteogenomics tell us?

- Locations of “Novel” Genes
- Confirm gene models
- Correct gene models
- Post-translational processing
Core Task of Proteogenomics

Raw MS/MS Data + Genome + Gene Models → Peptides Mapped to Genome
Example: Pear Scab

- Important crop pest
- Small genome (50 Mb)
- “Effector” proteins crucial to fungal invasion are small
Experiments
Results

• Confirmed ~1000 Predicted Genes (17 Effectors)

• “Discovered” 32 New Genes (13 Effectors)

• Confirmed 30 Post-translational Cleavages (22 Signal Peps, 2 Propeps, 6 Mitochondrial Targeting)

• 6 Gene Extensions, 490 Confirmed splice sites
Novel Genes in Trackster
Where to get the tools

• **Proteogenomics Workflow**
  
  Under Proteomics on the *Test* Toolshed
  proteogenomics_msms_to_gff_workflow

• **MS/MS Visualisations**
  
  https://bitbucket.org/Andrew_Brock/proteomics-visualise

• **Vagrant VM**
  
  https://bitbucket.org/iracooke/protk_vagrant_ubuntu_min

• **Proteomics Tools**
  
  https://github.com/iracooke/protk
Acknowledgements

★ Venturia Biology & Mass Spec
  Kim Plummer, Suresh Mathivanan, Pierre Faou, Daniel Jones, Vignesh Jayachandran

★ Galaxy Community
  Bjoern Groening, John Chilton

★ Funding
  VLSCI, LaTrobe e-Research, LaTrobe University