Galaxy in a proteomics core facility

Jeremy D Volkening1,*, Grzegorz Sabat1, Gregory A Barrett-Wilt1, and Michael R Sussman2

1Biotechnology Center, University of Wisconsin-Madison
2presenter: volkening@wisc.edu

Website: N/A

Repository: all relevant code will be made available on GitLab prior to the conference and referenced on the slides/poster.

Galaxy provides a mature and widely-adopted platform for accessible and reproducible data analysis in the life sciences. While the genomics community remains the largest user group, an increasing number of proteomics and “multi-omics” tools and workflows are being developed for use in Galaxy, primarily under the Galaxy-P umbrella. Consequently, Galaxy is nearing a state where it can compete with commercial platforms such as Proteome Discoverer (Thermo Scientific) for certain uses. In addition, the REST API provides a flexible way of interfacing Galaxy directly with mass spectrometers and other instruments. This enables a tight integration between data producers and consumers. Data can be automatically uploaded into Galaxy as it is generated, organized into data libraries, easily shared with clients, and automatically submitted to predefined workflows such as quality control reports and user-defined analyses.

We present here one such use case currently under testing in a mass spectrometry core facility. The infrastructure is described, including the instrument interface, cloud-based on-demand Pulsar instances for Windows-dependent file conversion, and REST-based interactions with the local Galaxy server. We outline the perceived benefits from moving to a Galaxy-centric analysis infrastructure, including increased automation and the ability to provide an unlimited number of “seats” on the Galaxy server for on-site and remote clients. We also describe current shortcomings that may prevent full adoption of such a system in production, particularly the dearth of built-in visualization tools for mass spectrometry data. These discussions incorporate end-user feedback relating to potential areas for improvement in usability and functionality compared with tools currently in use. Finally, we describe migration of an existing quality control pipeline to a Galaxy workflow and discuss the utility of automatically triggered workflows in a core facility environment.

Figure 1: Infrastructure for a Galaxy-based mass spectrometry data management and analysis platform