

MASCP Gator: An aggregating portal for proteomic data

H J. Joshi,¹ M Hirsch-Hoffmann,² K Bärenfaller,² W Gruissem,² S Baginsky,³ R Schmidt,⁴ W X. Shulze,⁴ Q Sun,⁵ K J. van Wijk,⁵ V Egelhofer,⁶ S Weinkoop,⁶ W Weckwerth,⁶ C Bruley,⁷ N Rolland,⁷ T Toyoda,⁸ H Nakagami,⁸ A M.E. Jones,⁹ S P. Briggs,¹⁰ I Castleden,¹¹ S K. Tanz,¹¹ A H. Millar and J L. Heazlewood¹

1: Joint BioEnergy Institute, Berkeley, California
2: Institute of Plant Sciences, ETH Zürich, Switzerland
3: Martin-Luther-University, Halle-Wittenberg
4: MPI for Molecular Plant Physiology, Potsdam-Golm, Germany
5: Department of Plant Biology, Cornell University, NY, USA
6: Molecular Systems Biology, University of Vienna, Vienna, Austria
7: CEA, France
8: RIKEN, Japan
9: The Sainsbury Laboratory, UK
10: UC San Diego, California
11: ARC Centre of Excellence in Plant Energy Biology, Uni. WA, Australia

Introduction

Bioinformatic resources often exist as islands of data on the internet. The ability to bring in multiple sources of data and comparatively examine them is hampered by the interfaces to access these data sources. The MASCP-Proteomics consortium presented a unique forum to undertake a collaborative bioinformatics project to tie together the data resources in one easy to access portal.

The Gator portal is a simple web page that allows for retrieval of proteomics data from across the MASCP consortium, and presents it in an easy to understand, summarised view. Through a client-side approach to aggregation, the data presented by the aggregator is always up-to-date.

Taking advantage of the agility of web technologies, all the components to make up the portal are available as open source libraries, so that this functionality can be built into not only the original resources, but other secondary resources such as wikis.

Methods

Development of the portal was a three step process: Identification of data resources and types, establishing the protocol, and building the infrastructure.

Identification of resources: Across the different services, a number of data types were identified - MS/MS peptide data, single amino acid modifications, and generalised localisation data.

Protocol establishment: As the resources being displayed are diverse, a versatile wire protocol (JSON) is used over HTTP to transfer data from the services. Server API definition was left loose to aid with adoption of the service, while API strictness was enforced on the client.

Infrastructure building: The client was built using HTML, JavaScript and SVG, using the AJAX technology to retrieve data from the servers.

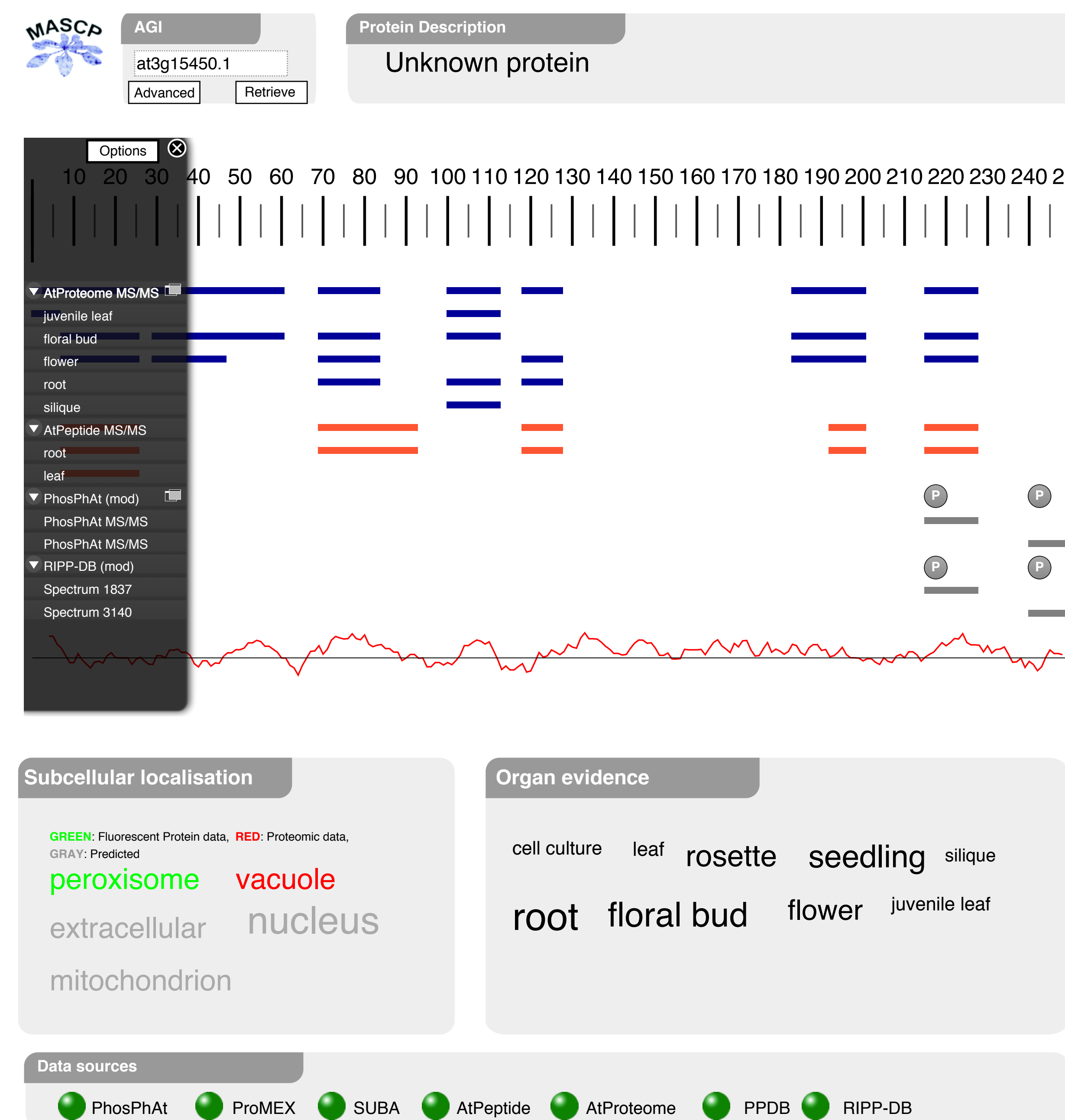


Figure 1: Screenshot of GATOR tool for unknown protein AT3G15450.1

Results

The portal was built and hosted at <http://gator.masc-proteomics.org>. By inserting a single TAIR Arabidopsis Genome Initiative (AGI) identifier, an overview of the data can be easily presented. AT3G15450.1 is an unknown protein, that has had mass spectrometric, localisation and tissue expression data collected about it (Figure 1).

Two phosphopeptides have been recorded by PhosPhAt, at positions 216-228 and 241-253, with the 216-228 peptide also observed without a phosphorylation site by AtProteome and ATPeptide. This is interesting, as it demonstrates that the phosphorylation site on the 216-228 phosphopeptide is being actively regulated.

Using the tool, experiments can be readily designed for single reaction monitoring (SRM) experiments to quantitate changes in a protein abundance. The MASCP-Gator clearly indicates peptides that are compatible with proteomics-based mass spectrometry and are readily detected by these instruments. Peptides consistently identified with no miss-cleave or over-

Included data sets

AtProteome
PhosPhAt
SUBA
AtPeptide
ProMeX
TAIR
PPDB
RIPP-DB
AT_CHLORO

Features

Zoom-sensitive display of data
MS/MS peptide data
Localisation data
Tissue data
Hydropathy plot

lap can be readily selected, such as GPYPADQVLR at position 120 amino acids (Figure 1), and used for SRM-based quantification with confidence. Moreover measuring changes in the phosphorylation state of AT3G15450.1 could be easily accomplished by selecting using both the phosphorylated and unphosphorylated peptide at position 216-228 and examining phosphorylation changes by employing SRM analysis of various tissues (e.g. flower to leaf). Using the expression information, the choice of tissues to study can be directed.

Conclusions

The aggregating portal presented here provides a novel tool to the proteomics community, allowing scientists to better design experiments, and through its visual nature understand the nature of their protein of interest at a glance. Further improvements to this tool are planned to allow the comparative display of protein data, enriching the functionality.

Acknowledgements

This work was developed out of the Multinational Arabidopsis Steering Committee, Proteomics Subcommittee and represents a coordination of efforts between a variety of international research institutions, organizations and funding agencies.

<http://gator.masc-proteomics.org/>