



MASCP gator: an overview of the Arabidopsis proteomic aggregation portal

Gregory W. Mann¹, Paul C. Calley¹, Hiren J. Joshi² and Joshua L. Heazlewood^{1*}

¹ Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

² Copenhagen Center for Glycomics, Institute for Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

Edited by:

Hirofumi Nakagami, RIKEN, Japan

Reviewed by:

Hans-Peter Braun, Hannover University, Germany

Martin Hajduch, Slovak Academy of Sciences, Slovakia

*Correspondence:

Joshua L. Heazlewood, Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, One Cyclotron Road MS 978-4466, Berkeley, CA 94720, USA
e-mail: jlheazlewood@lbl.gov

A key challenge in the area of bioinformatics in the coming decades is the ability to manage the wealth of information that is being generated from the variety of high throughput methodologies currently being undertaken in laboratories across the world. While these approaches have made available large volumes of data to the research community, less attention has been given to the problem of how to intuitively present the data to enable greater biological insights. Recently, an attempt was made to tackle this problem in the area of Arabidopsis proteomics. The model plant has been the target of countless proteomics surveys producing an exhaustive array of data and online repositories. The MASCP Gator is an aggregation portal for proteomic data currently being produced by the community and unites a large collection of specialized resources to a single portal (<http://gator.masc-proteomics.org/>). Here we describe the latest additions, upgrades and features to this resource further expanding its role into protein modifications and genome sequence variations.

Keywords: proteomics, Arabidopsis, mass spectrometry, database, protein modifications, single nucleotide polymorphisms

BACKGROUND

The utilization of mass spectrometry for the identification of proteins from plant samples has been widely applied for the past decade (Heazlewood, 2011). The technology has been widely applied in large scale proteomic surveys of the model plant Arabidopsis (Castellana et al., 2008; Baerenfaller et al., 2011). These studies in Arabidopsis have resulted in the production of large quantities of data which are subsequently available through a variety of online portals and resources (Weckwerth et al., 2008). The construction of these boutique databases can result in a fragmentation of knowledge in a given biological system. In an attempt to mitigate this problem, the proteomics subcommittee of the Multinational Arabidopsis Steering Committee (MASCP) developed a data aggregation portal (MASCP Gator) to display proteomic data to the community (Joshi et al., 2011).

The MASCP Gator portal was originally made available in 2011 (<http://gator.masc-proteomics.org/>), and during the intervening period, the tool has been extensively updated, improving functionality and usability. A considerable amount of data and information have been added and the core sequences have been updated to the latest release available from The Arabidopsis Information Resource (TAIR) (Lamesch et al., 2012). Many improvements to the user interface have been made, making the tool easier to use, and exposing more of the tool for use in scripting environments. This article describes updates to the portal highlighting new features and summarizing the current state of protein identifications in the model plant Arabidopsis.

IMPLEMENTATION OF THE MASCP GATOR

The original concept of the MASCP Gator was to provide a visual representation of distributed proteomics data that is both

comprehensive and simple to use. The idea to connect a collection of boutique biological databases through a series of web services was not a new concept (Wilkinson and Links, 2002). The MASCP Gator was originally implemented using HTML, JavaScript and scalable vector graphics with a server-side component for caching and historical tracking of data written in JavaScript and the Node.JS runtime environment (Joshi et al., 2011). Web services were provided from the data sources in JSON format responding to a query based upon a given Arabidopsis Gene Identifier (AGI). The source code is available online at <http://gator.masc-proteomics.org/source/> as well as documentation, unit tests, and examples so that individuals can utilize the libraries developed for the aggregator.

The initial construction of the MASCP Gator (Joshi et al., 2011) comprised aggregated information from SUBA (Heazlewood et al., 2007), AtProteome (Baerenfaller et al., 2008), ProMEX (Hummel et al., 2007; Wienkoop et al., 2012), PhosPhAt (Heazlewood et al., 2008; Durek et al., 2010; Arsova and Schulze, 2012), PPDB (Sun et al., 2009), RIPP-DB (Nakagami et al., 2010), AT_CHLORO (Ferro et al., 2010; Bruley et al., 2012), and AtPeptide (Castellana et al., 2008) and included underlying protein information and descriptions from TAIR (release 9) (Swarbreck et al., 2008). Since the initial release of the MASCP Gator a number of updates and changes have been undertaken. These have included the recent evolution of the AtProteome resource to pep2pro which coincided with updates to their analysis pipeline and the addition of new data (Baerenfaller et al., 2011; Hirsch-Hoffmann et al., 2012). Upgrades to the underlying protein models and descriptions have also been undertaken to incorporate TAIR10 (Lamesch et al., 2012).

A screenshot of the current release interface for the MASCP Gator is shown in **Figure 1**. There are a number of new features and additions to the interface which will be outlined below.

RECENT UPDATES TO THE MASCP GATOR

The MASCP Gator was designed in a modular manner to enable the incorporation of additional data sources as they are developed by the community. Moreover, the MASCP Gator's adoption of the NodeJS framework allows the rapid development of additional functionality including the introduction of static data sources.

A number of dynamic data sources have been added to the MASCP Gator since its initial development and release (Joshi et al., 2011). The Plant Protein Phosphorylation DataBase (P³DB) initially focused on providing protein phosphorylation data from

a variety of plant species, with a major focus on *Brassica napus* (Gao et al., 2009). More recently the P³DB resource has begun housing experimental phosphorylation data from *Arabidopsis thaliana* as well as a range of other species including rice, Medicago, and soybean (Yao et al., 2012). Although much of the Arabidopsis data from P³DB overlaps considerably with existing phosphorylation resources already integrated with the MASCP Gator (e.g., PhosPhAt and RIPP-DB), redundancy is a useful mechanism to offset data stagnation from resource providers. The GelMap resource was also recently included and houses a range of two-dimensional gel electrophoresis images from plant extracts with mapped protein identifications from proteomic surveys (Rode et al., 2011). Its initial focus has been organelle proteomes from Arabidopsis, namely mitochondria and plastid,

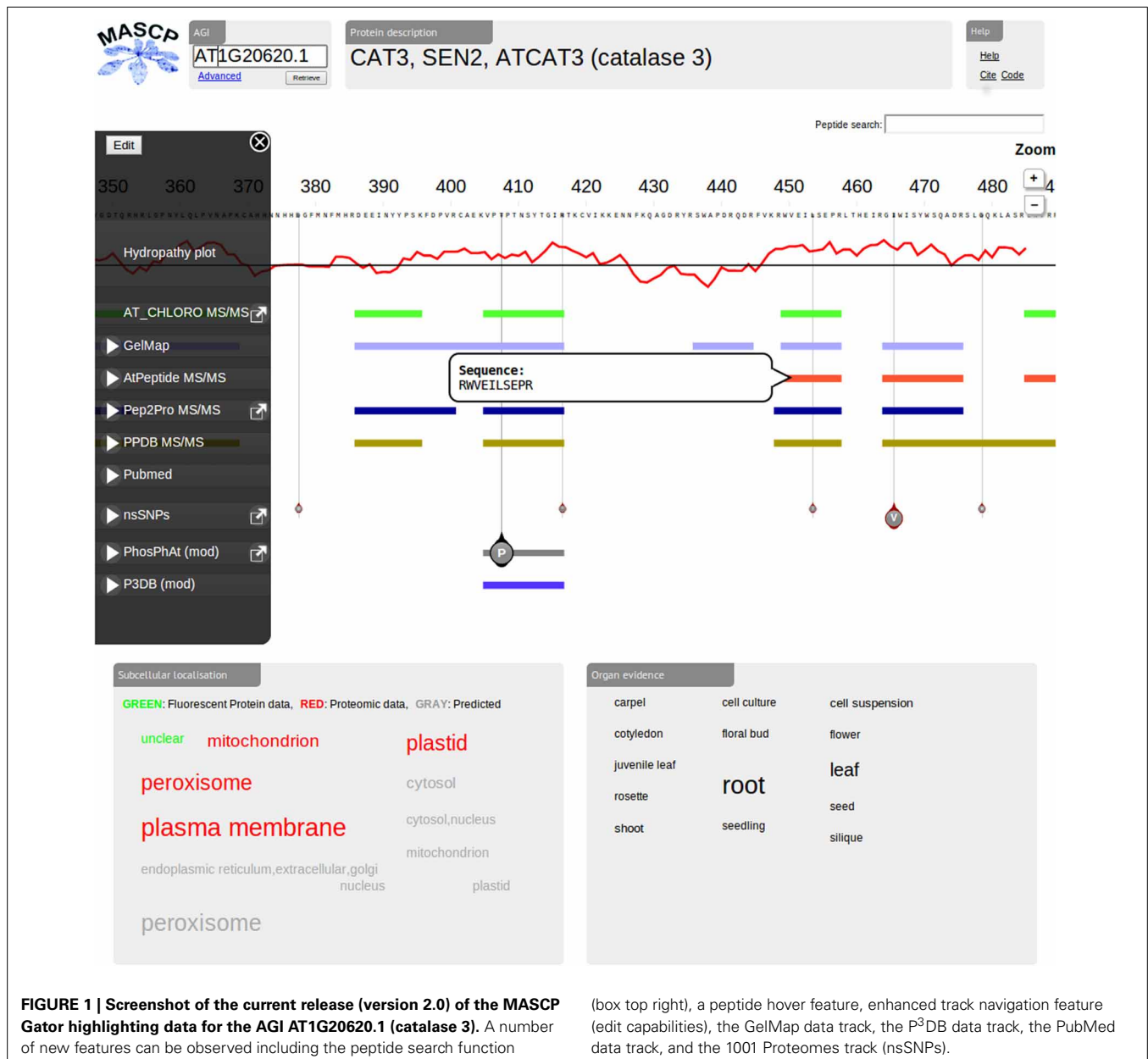


FIGURE 1 | Screenshot of the current release (version 2.0) of the MASCP Gator highlighting data for the AGI AT1G20620.1 (catalase 3). A number of new features can be observed including the peptide search function

(box top right), a peptide hover feature, enhanced track navigation feature (edit capabilities), the GelMap data track, the P³DB data track, the PubMed data track, and the 1001 Proteomes track (nsSNPs).

but this will likely expand as the community is encouraged to upload published data at this repository (Senkler and Braun, 2012). The inclusion of these resources brings the total number of peptides displayed by the MASCP Gator to over 10 million, which is comprised of around 220,000 unique peptides (Mann et al., 2013). The MASCP Gator displays information (experimental and non-experimental) for all 35,386 proteins currently available in TAIR10, which includes nearly 8000 proteins derived from alternate splicing models (designated as AGI code plus .2, .3, .4 and etc). Protein information will be displayed in the MASCP Gator at the alternate splice level if specified by a particular resource or publication. Where data only indicate information at the locus level, data is displayed for the representative gene model (generally the .1 version). In total, there is proteomic information for 24,811 Arabidopsis proteins covering over 70% of the potential proteome when compared to all proteins derived from alternate gene models (Table 1). The majority of the nearly 10,600 proteins with no experimental information mainly consist of alternate gene models (data not shown). A cursory examination of peptide coverage of the 24,811 proteins with experimental information shows that over 50% of the proteins have greater than 20% coverage (Figure S1). This simple analysis highlights the utility of data aggregation as these numbers suggest very good experimental evidence for nearly 12,500 Arabidopsis proteins.

A collection of static data related to protein features in Arabidopsis has also been included in the MASCP Gator. These include the addition of a PubMed track to capture published proteomics data not currently contained in online repositories (Roberts, 2001). The original publication of data displayed on this track can be accessed through a hyperlinked PubMed identifier (PMID). This feature will enable the collation and display of ongoing proteomics data in the model plant Arabidopsis. In the initial release of the MASCP Gator the only posttranslational modification (PTM) displayed was phosphorylation. More recently we have endeavored to add and annotate an expanded array of PTMs that have been characterized in the literature and

currently not readily accessible. The following have been recently added to the MASCP Gator as discrete tracks for visualization:

1. Ubiquitination, a covalent attachment of ubiquitin that typically occurs to lysine residues of proteins. These modifications can affect protein interactions and targeting as well as signal their degradation via the proteasome (Vierstra, 2009). The MASCP Gator currently displays 317 experimentally determined ubiquitination sites from 214 distinct Arabidopsis proteins defined from four studies (Maor et al., 2007; Saracco et al., 2009; Book et al., 2010; Kim et al., 2013). These modification sites are labeled with UBQ on a dedicated track.
2. Protein glycosylation represents one of the most complex PTMs in protein biochemistry and affects protein folding, stability, function and targeting (Gomord et al., 2010). In plants, few studies have attempted to extensively characterize this modification. Recently, a large-scale analysis of N-linked glycosylation was undertaken with nearly 2200 sites identified in Arabidopsis (Zielinska et al., 2012). These data have a dedicated glycosylation track labeled with CHO within the interface.
3. Methionine oxidation is generally regarded as a non-biological modification of proteins that occurs during sample preparation. While the oxidation of methionine residues as a consequence of cellular oxidative damage would provide a biological source of this PTM, its role in cellular signaling is also being explored (Hardin et al., 2009). A recent proteomic profile of methionine oxidation after the addition of a cGMP analog resulted in the identification of nearly 500 induced sites from nearly 400 proteins (Maroundedze et al., 2013). These modifications are contained on a dedicated track and each site labeled MOX.
4. The N-terminal processing of targeting signals or presequences is a common feature of proteins imported into the mitochondrion or plastid. This processed polypeptide represents the mature form of the protein. A number of studies have characterized the N-terminal of these mature organelle proteins using proteomic approaches (Zybailov et al., 2008; Huang et al., 2009); while various studies have also defined these mature N-terminal sites by Edman degradation (Prime et al., 2000; Kruff et al., 2001; Eubel et al., 2003; Millar et al., 2004). Collectively, over 130 proteins have had their N-terminal pre-sequence defined. The first amino acid is labeled with "MAT" in the MASCP Gator to indicate the start of the mature protein.
5. RNA editing is a process that results in posttranscriptional changes to mRNA that can often lead to changes to the resultant protein sequence. RNA editing in plants results in the conversion of cytidine to uridine and has thus been characterized in transcripts from mitochondria and plastids (Takenaka et al., 2008). A total of 478 RNA editing sites from fifty plastid and mitochondrial encoded genes that result in changes to the amino acid sequence are now available (Giege and Brennicke, 1999; Chateigner-Boutin and Small, 2007; Bentolila et al., 2008; Zehrmann et al., 2008). These changes are displayed with the substituted residue in a dedicated track RNA editing track.

Table 1 | Summary of proteomic derived information (August 2013) currently aggregated and displayed by the MASCP Gator.

Resource	Peptides	Proteins
SUBA	N/A	9319 ¹
pep2pro	6,178,840	21,451
PPDB	1,787,614	6905
ProMEX	695	248
AtPeptide	2,520,272	16,769
PhosPhAt	61,795	5486
AT_CHLORO	122,488	1263
RIPP-DB	5562	2244
P ³ DB	7575	3851
GelMap	21588	986
PubMed	21,774	2867

¹The protein value for SUBA corresponds to proteins with experimental localization data (MS or FP).

6. S-Nitrosylation is the reversible covalent attachment of nitric oxide to the thiol groups of cysteine residues. Nitric oxide appears to be an important signaling molecule in plant biology with an involvement in processes such as seed germination and floral senescence while also having a role in abiotic and biotic stresses. Its role in signaling appears to be directly linked to the modification of target proteins. Few studies have attempted to characterize the actual modified cysteine residues, and currently only about 60 sites have been mapped in *Arabidopsis* (Dixon et al., 2005; Palmieri et al., 2010; Fares et al., 2011). These modifications are labeled with SNO in the interface on a dedicated track.

In addition to these newly added proteomic-based resources, data from the *Arabidopsis* 1001 Genomes Project have also been incorporated (Weigel and Mott, 2009). These data represent non-synonymous single nucleotide polymorphisms (nsSNPs) resulting from the re-sequencing of *Arabidopsis* accessions (Joshi et al., 2012). The data is presented as amino acid substitutions which are scored according to their frequency of observation in each accession, relative to *Arabidopsis thaliana* (Col-0).

ENHANCEMENTS TO THE MASCP GATOR

A number of enhancements have been made to both the user interface and to server side components. Mobile devices are taking an increasing share of the personal computing market, and as such it is useful to develop web applications so that they can work well on these devices. Further, as this is the direction that the computing space is heading, it is prudent to build applications so that they support this computing paradigm natively. Two challenges need to be addressed when retro-fitting an application to support this usage model; firstly, the computational components need to be optimized so that the application is responsive on a lower powered CPU as found on these devices. The second challenge faced is that the user interface needs to be updated so that it does not rely upon a mouse for interaction. The MASCP Gator has been updated to support both these challenges and as a result is more responsive on both mobile and desktop devices. To further improve the utility and speed of the MASCP Gator a server-side component has been established. Previously, it was necessary to set up a proxy server to allow access to certain sites that had not yet fully implemented the data interchange interface. In addition, software was required to allow for hosting of data sets that were retrieved from the literature. These two pieces of functionality were consolidated to enable periodic crawling and the update of data sources from the data providers.

The user interface has now been enriched to allow for the reloading of individual data types. Since the portal employs client side caching, data from a prior visit may be outdated. The Data sources component of the display now provides a timestamp to indicate the age of the displayed data. If necessary, individual data sources can be reloaded by clicking the circular arrow associated with each data type. A pop-up window has been added to the peptide tracks that displays the peptide sequence when the user hovers the mouse over an experimentally identified peptide

on a track. This peptide sequence information can be copied using the mouse. A peptide search feature has also been added to the interface to enable a user to locate a peptide sequence in the currently displayed protein. If found, the peptide will be highlighted through the data tracks to provide context. The data tracks can now be manipulated and removed from the display. An “Edit” button on the track navigation section enables data track manipulation. Once activated, clicking the “X” adjacent to a given track will remove it from view. The track can be retrieved by refreshing the web page. The order of tracks can be manipulated by dragging and moving them while in edit mode. These features enable a better user experience with the resource.

FUTURE DEVELOPMENT

The incorporation and aggregation of new proteomic resources will be a primary driver of the MASCP Gator’s future development. These data will likely comprise supplementary material from published studies and the integration of new independent databases. While minor modifications to the user interface will occur, the ongoing collection and visual display of proteomics data is the primary function of this resource. With the ongoing development of the *Arabidopsis* Information Portal (International *Arabidopsis* Informatics, 2012) by the *Arabidopsis* Informatics Consortium it is likely that much of the functionality present in the MASCP Gator will migrate to this new portal. The development of services by the *Arabidopsis* proteomics community that have enabled the construction of the MASCP Gator will provide a seamless transition to the *Arabidopsis* Information Portal in the near future.

AUTHOR CONTRIBUTIONS

The manuscript was devised and written by Gregory W. Mann, Hiren J. Joshi, and Joshua L. Heazlewood. Data was collected and figures were created by Gregory W. Mann. Data updates for this implementation of the MASCP Gator were curated and added by Gregory W. Mann and Paul C. Calley. The core functionality of the MASCP Gator was developed by Hiren J. Joshi.

ACKNOWLEDGMENTS

The MASCP Gator was created with the support of members of the *Arabidopsis* proteomics community who assisted in the development of database specific services to enable data exchange. This work was conducted by the Joint BioEnergy Institute and is supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2013.00411/abstract>

Figure S1 | Histogram of protein coverage (%) for the 24,811 proteins with peptide data currently displayed in the MASCP Gator. Greater than 50% of identified proteins have over 20% protein coverage.

REFERENCES

- Arsova, B., and Schulze, W. X. (2012). Current status of the plant phosphorylation site database PhosPhAt and its use as a resource for molecular plant physiology. *Front. Plant Sci.* 3:132. doi: 10.3389/fpls.2012.00132
- Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., et al. (2008). Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 320, 938–941. doi: 10.1126/science.1157956
- Baerenfaller, K., Hirsch-Hoffmann, M., Svozil, J., Hull, R., Russenberger, D., Bischof, S., et al. (2011). pep2pro: a new tool for comprehensive proteome data analysis to reveal information about organ-specific proteomes in *Arabidopsis thaliana*. *Integr. Biol.* 3, 225–237. doi: 10.1039/c0ib00078g
- Bentolila, S., Elliott, L. E., and Hanson, M. R. (2008). Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics* 178, 1693–1708. doi: 10.1534/genetics.107.073585
- Book, A. J., Gladman, N. P., Lee, S. S., Scaif, M., Smith, L. M., and Vierstra, R. D. (2010). Affinity purification of the *Arabidopsis* 26S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* 285, 25554–25569. doi: 10.1074/jbc.M110.136622
- Bruley, C., Dupieris, V., Salvi, D., Rolland, N., and Ferro, M. (2012). AT_CHLORO: a chloroplast protein database dedicated to subplastidial localization. *Front. Plant Sci.* 3:205. doi: 10.3389/fpls.2012.00205
- Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M., Bafna, V., and Briggs, S. P. (2008). Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 21034–21038. doi: 10.1073/pnas.0811066106
- Chateigner-Boutin, A. L., and Small, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* 35, e114. doi: 10.1093/nar/gkm640
- Dixon, D. P., Skipsey, M., Grundy, N. M., and Edwards, R. (2005). Stress-induced protein S-glutathionylation in *Arabidopsis*. *Plant Physiol.* 138, 2233–2244. doi: 10.1104/pp.104.058917
- Durek, P., Schmidt, R., Heazlewood, J. L., Jones, A., Maclean, D., Nagel, A., et al. (2010). PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res.* 38, D828–D834. doi: 10.1093/nar/gkp810
- Eubel, H., Jansch, L., and Braun, H. P. (2003). New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. *Plant Physiol.* 133, 274–286. doi: 10.1104/pp.103.024620
- Fares, A., Rossignol, M., and Peltier, J. B. (2011). Proteomics investigation of endogenous S-nitrosylation in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 416, 331–336. doi: 10.1016/j.bbrc.2011.11.036
- Ferro, M., Brugièrè, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., et al. (2010). AT_CHLORO, A comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* 9, 1063–1084. doi: 10.1074/mcp.M900325-MCP200
- Gao, J., Agrawal, G. K., Thelen, J. J., and Xu, D. (2009). P3DB: a plant protein phosphorylation database. *Nucleic Acids Res.* 37, D960–D962. doi: 10.1093/nar/gkn733
- Giege, P., and Brennicke, A. (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. U.S.A.* 96, 15324–15329. doi: 10.1073/pnas.96.26.15324
- Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., et al. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x
- Hardin, S. C., Larue, C. T., Oh, M. H., Jain, V., and Huber, S. C. (2009). Coupling oxidative signals to protein phosphorylation via methionine oxidation in *Arabidopsis*. *Biochem. J.* 422, 305–312. doi: 10.1042/BJ20090764
- Heazlewood, J. L. (2011). The Green proteome: challenges in plant proteomics. *Front. Plant Sci.* 2:6. doi: 10.3389/fpls.2011.00006
- Heazlewood, J. L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D., et al. (2008). PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res.* 36, D1015–D1021. doi: 10.1093/nar/gkm812
- Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., and Millar, A. H. (2007). SUBA: the *Arabidopsis* subcellular database. *Nucleic Acids Res.* 35, D213–D218. doi: 10.1093/nar/gkl863
- Hirsch-Hoffmann, M., Gruissem, W., and Baerenfaller, K. (2012). pep2pro: the high-throughput proteomics data processing, analysis, and visualization tool. *Front. Plant Sci.* 3:123. doi: 10.3389/fpls.2012.00123
- Huang, S., Taylor, N. L., Whelan, J., and Millar, A. H. (2009). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol.* 150, 1272–1285. doi: 10.1104/pp.109.137885
- Hummel, J., Niemann, M., Wienkoop, S., Schulze, W., Steinhauser, D., Selbig, J., et al. (2007). ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics* 8:216. doi: 10.1186/1471-2105-8-216
- International Arabidopsis Informatics, C. (2012). Taking the next step: building an Arabidopsis information portal. *Plant Cell* 24, 2248–2256. doi: 10.1105/tpc.112.100669
- Joshi, H. J., Christiansen, K. M., Fitz, J., Cao, J., Lipzen, A., Martin, J., et al. (2012). 1001 Proteomes: a functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics* 28, 1303–1306. doi: 10.1093/bioinformatics/bts133
- Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., et al. (2011). MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *Plant Physiol.* 155, 259–270. doi: 10.1104/pp.110.168195
- Kim, D. Y., Scaif, M., Smith, L. M., and Vierstra, R. D. (2013). Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in *Arabidopsis*. *Plant Cell* 25, 1523–1540. doi: 10.1105/tpc.112.108613
- Krufft, V., Eubel, H., Jansch, L., Werhahn, W., and Braun, H. P. (2001). Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis*. *Plant Physiol.* 127, 1694–1710. doi: 10.1104/pp.010474
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Mann, G. W., Joshi, H. J., Petzold, C. J., and Heazlewood, J. L. (2013). Proteome coverage of the model plant *Arabidopsis thaliana*: implications for shotgun proteomic studies. *J. Proteomics* 79, 195–199. doi: 10.1016/j.jprot.2012.12.009
- Maor, R., Jones, A., Nuhse, T. S., Studholme, D. J., Peck, S. C., and Shirasu, K. (2007). Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics* 6, 601–610. doi: 10.1074/mcp.M600408-MCP200
- Marondedze, C., Turek, I., Parrott, B., Thomas, L., Jankovic, B., Lilley, K. S., et al. (2013). Structural and functional characteristics of cGMP-dependent methionine oxidation in *Arabidopsis thaliana* proteins. *Cell Commun. Signal.* 11, 1. doi: 10.1186/1478-811X-11-1
- Millar, A. H., Eubel, H., Jansch, L., Krufft, V., Heazlewood, J. L., and Braun, H. P. (2004). Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant specific subunits. *Plant Mol. Biol.* 56, 77–90. doi: 10.1007/s11103-004-2316-2
- Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., et al. (2010). Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* 153, 1161–1174. doi: 10.1104/pp.110.157347
- Palmieri, M. C., Lindermayr, C., Bauwe, H., Steinhauser, C., and Durner, J. (2010). Regulation of plant glycine decarboxylase by S-nitrosylation and glutathionylation. *Plant Physiol.* 152, 1514–1528. doi: 10.1104/pp.109.152579
- Prime, T. A., Sherrier, D. J., Mahon, P., Packman, L. C., and Dupree, P. (2000). A proteomic analysis of organelles from *Arabidopsis thaliana*. *Electrophoresis* 21, 3488–3499. doi: 10.1002/1522-2683(20001001)21:16<3488::AID-ELPS3488>3.0.CO;2-3
- Roberts, R. J. (2001). PubMed central: the GenBank of the published literature. *Proc. Natl. Acad. Sci. U.S.A.* 98, 381–382. doi: 10.1073/pnas.98.2.381
- Rode, C., Senkler, M., Klodmann, J., Winkelmann, T., and Braun, H. P. (2011). GelMap-a novel software tool for building and presenting proteome reference maps. *J. Proteomics* 74, 2214–2219. doi: 10.1016/j.jprot.2011.06.017

- Saracco, S. A., Hansson, M., Scalf, M., Walker, J. M., Smith, L. M., and Vierstra, R. D. (2009). Tandem affinity purification and mass spectrometric analysis of ubiquitylated proteins in *Arabidopsis*. *Plant J.* 59, 344–358. doi: 10.1111/j.1365-313X.2009.03862.x
- Senkler, M., and Braun, H. P. (2012). Functional annotation of 2D protein maps: the GelMap portal. *Front. Plant Sci.* 3:87. doi: 10.3389/fpls.2012.00087
- Sun, Q., Zybailov, B., Majeran, W., Friso, G., Olinares, P. D., and Van Wijk, K. J. (2009). PPDB, The plant proteomics database at Cornell. *Nucleic Acids Res.* 37, D969–D974. doi: 10.1093/nar/gkn654
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., et al. (2008). The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36, D1009–D1014. doi: 10.1093/nar/gkm965
- Takenaka, M., Verbitskiy, D., Van Der Merwe, J. A., Zehrmann, A., and Brennicke, A. (2008). The process of RNA editing in plant mitochondria. *Mitochondrion* 8, 35–46. doi: 10.1016/j.mito.2007.09.004
- Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10, 385–397. doi: 10.1038/nrm2688
- Weckwerth, W., Baginsky, S., Van Wijk, K., Heazlewood, J. L., and Millar, A. H. (2008). The Multinational *Arabidopsis* Steering Subcommittee for Proteomics assembles the largest proteome database resource for plant systems biology. *J. Proteome Res.* 7, 4209–4210. doi: 10.1021/pr800480u
- Weigel, D., and Mott, R. (2009). The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biol.* 10, 107. doi: 10.1186/gb-2009-10-5-107
- Wienkoop, S., Staudinger, C., Hoehenwarter, W., Weckwerth, W., and Egelhofer, V. (2012). ProMEX—a mass spectral reference database for plant proteomics. *Front. Plant Sci.* 3:125. doi: 10.3389/fpls.2012.00125
- Wilkinson, M. D., and Links, M. (2002). BioMOBY: an open source biological web services proposal. *Brief Bioinform.* 3, 331–341. doi: 10.1093/bib/3.4.331
- Yao, Q., Bollinger, C., Gao, J., Xu, D., and Thelen, J. J. (2012). P3DB: an integrated database for plant protein phosphorylation. *Front. Plant Sci.* 3:206. doi: 10.3389/fpls.2012.00206
- Zehrmann, A., Van Der Merwe, J. A., Verbitskiy, D., Brennicke, A., and Takenaka, M. (2008). Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of *Arabidopsis thaliana*. *Mitochondrion* 8, 319–327. doi: 10.1016/j.mito.2008.07.003
- Zielinska, D. E., Gnad, F., Schropp, K., Wisniewski, J. R., and Mann, M. (2012). Mapping *N*-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* 46, 542–548. doi: 10.1016/j.molcel.2012.04.031
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., et al. (2008). Sorting signals, *N*-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3:e1994. doi: 10.1371/journal.pone.0001994

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 September 2013; accepted: 27 September 2013; published online: 23 October 2013.

Citation: Mann GW, Calley PC, Joshi HJ and Heazlewood JL (2013) MASCP gator: an overview of the *Arabidopsis* proteomic aggregation portal. *Front. Plant Sci.* 4:411. doi: 10.3389/fpls.2013.00411

This article was submitted to *Plant Proteomics*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2013 Mann, Calley, Joshi and Heazlewood. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.