DOI 10.1002/pmic.201000256 *Proteomics* 2011, *11*, 2354–2358

DATASET BRIEF

Proteomic characterization of the major arthropod associates of the carnivorous pitcher plant *Sarracenia purpurea*

Nicholas J. Gotelli[1], *Aidan M. Smith*[1], *Aaron M. Ellison*[2] *and Bryan A. Ballif*[1,3]

[1] Department of Biology, University of Vermont, Burlington, VT, USA
[2] Harvard Forest, Harvard University, Petersham, MA, USA
[3] Vermont Genetics Network Proteomics Facility, University of Vermont, Burlington, VT, USA

Received: April 19, 2010
Revised: January 10, 2011
Accepted: February 28, 2011

The array of biomolecules generated by a functioning ecosystem represents both a potential resource for sustainable harvest and a potential indicator of ecosystem health and function. The cupped leaves of the carnivorous pitcher plant, *Sarracenia purpurea*, harbor a dynamic food web of aquatic invertebrates in a fully functional miniature ecosystem. The energetic base of this food web consists of insect prey, which is shredded by aquatic invertebrates and decomposed by microbes. Biomolecules and metabolites produced by this food web are actively exchanged with the photosynthesizing plant. In this report, we provide the first proteomic characterization of the sacrophagid fly (*Fletcherimyia fletcheri*), the pitcher plant mosquito (*Wyeomyia smithii*), and the pitcher-plant midge (*Metriocnemus knabi*). These three arthropods act as predators, filter feeders, and shredders at distinct trophic levels within the *S. purpurea* food web. More than 50 proteins from each species were identified, ten of which were predominantly or uniquely found in one species. Furthermore, 19 peptides unique to one of the three species were identified using an assembled database of 100 metazoan myosin heavy chain orthologs. These molecular signatures may be useful in species monitoring within heterogeneous ecosystem biomass and may also serve as indicators of ecosystem state.

Keywords:
Animal proteomics / Ecosystem / LC-MS/MS / Nonmodel organism / Pitcher Plant

A major research focus in basic and applied ecology is to understand the origin and maintenance of ecosystem services [1]. Ecosystem services include regulatory services such as carbon sequestration, climate control, and decomposition. Ecosystems also provide services in the form of plant and animal products that can be sustainably harvested [2]. Furthermore, plant and animal products are broken down to yield a suite of elemental nutrients and more complex organic molecules, including proteins. However, the generation of even a simple inventory of molecular ecosystem services within an intact food web is a daunting prospect. The physical and biological boundaries of a food web are usually impossible to objectively define, and the spatial extent of a typical natural food web is simply too large for a meaningful comprehensive study of biomolecular services. Until recently, it has not been possible for ecologists to even begin an efficient survey of the full spectrum of molecular products from an intact ecosystem [3]. As a consequence, most studies of the biochemical throughput of ecosystems have been limited to analyses of a few critical nutrients and stoichiometric ratios [4]. In this study, we report an initial proteomic characterization of a naturally intact and spatially tractable ecosystem [5]. Our proteomic survey revealed protein- and peptide-based biomarkers that could be used to monitor specific taxa or particular proteins within a defined food web.

An entire aquatic food web (Fig. 1A) is harbored within the cupped leaves of the carnivorous pitcher plant *Sarracenia purpurea*. The leaves of this plant form a simple, open tube that fills with roughly 10 mL of rainwater and functions as a

Correspondence: Dr. Bryan A. Ballif, Department of Biology, University of Vermont, 120A Marsh Life Science Building, 109 Carrigan Drive, Burlington, VT 05405, USA
E-mail: bballif@uvm.edu
Fax: +1-802-656-2914

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

trap that catches invertebrate prey (mostly ants and flies [6]). The dead carcasses of captured prey are the base of a distinctive aquatic food web that includes microbes, protozoa, rotifers, mites, and the larval stages of three dipterans: midges, mosquitoes, and sarcophagid flies. The midge and sarcophagid larvae shred the prey, which is further decomposed by the microbial component of the food web. Protists, mites, and mosquito larvae feed on the bacteria, and late instar sarcophagid larvae feed on protists and mosquito larvae [7]. Nutrients and biomolecules released by this decomposition process are actively sequestered by the living plant, which in turn releases oxygen and probably other metabolites and organic molecules to the pitcher liquid [8].

A major goal is to comprehensively identify the protein component of the biomolecular services provided by and exchanged within this ''*Sarracenia* microecosystem'' and to identify quantifiable protein signatures indicative of its biological state. Toward this goal, we first sought to conduct a proteomic survey to determine the protein signatures of the aquatic larvae of the three common dipteran species residing within the pitcher plant: a sacrophagid fly (*Fletcherimyia fletcheri*), the pitcher plant mosquito (*Wyeomyia smithii*), and the pitcher-plant midge (*Metriocnemus knabi*). These three species occur at the base, the middle, and the top of this aquatic food web, and they are commonly collected in the leaves of *S. purpurea* throughout its extensive geographic range [9]. We reasoned that each of these species might display a unique protein profile that would help us to identify the origin of particular proteins encountered during more comprehensive surveys of the pitcher plant liquid and which may be indicators of ecosystem state. For more information on sample collection and preparation, see Supporting Information Materials and methods.

Figure 1B shows the protein profiles obtained from the isolated resident arthropods (left panel) and three samples of mixed particulates from individual pitcher plants that represent medium, low and high biomass density respectively (right panel). Although the profiles of the isolated species show a number of visual similarities, characteristic differences can be observed in relative band intensities, creating a SDS-PAGE ''signature'' of each species. In some cases, the SDS-PAGE signature of a given isolate suggested a possible diagnostic for determining the dominance of a given larval species within a mixed-species sample (compare the SDS-PAGE signature of *W. smithii* in lane 1 of the left panel of Fig. 1B with mixed particulate 1 (P1) in the right panel). However, as in the case of particulate fraction 3 (P3), when the pitcher plant had an abundance of captured arthropod prey, no one of the larval signatures was readily discernable. Together, these data suggest a rough diagnostic for determining protein signatures of the three major pitcher plant dipteran species in isolation or when they dominate the biomass of a given particulate fraction.

Although the SDS-PAGE profiles might prove useful, identifying species-specific, or at least species-predominant, proteins would add an additional level of confidence in establishing protein signatures for future proteomic surveys of the pitcher plant food web. We therefore endeavored to catalogue the major proteins of the three dipteran larvae. Such an analysis would also provide the first proteomic reference database of these organisms. The entire gel lane for each organism shown in Fig. 1B (left panel) was divided into 16 (*F. fletcheri*) or 17 slices. Major bands were kept common to facilitate comparisons. Gel slices were diced into 1 mm cubes and subjected to in-gel tryptic digestion and peptide extraction as described previously [10]. MS and data analysis is described in the Supporting Information Materials and methods.

More than 50 proteins from each species were identified following SEQUEST search analysis using the *Drosophila*

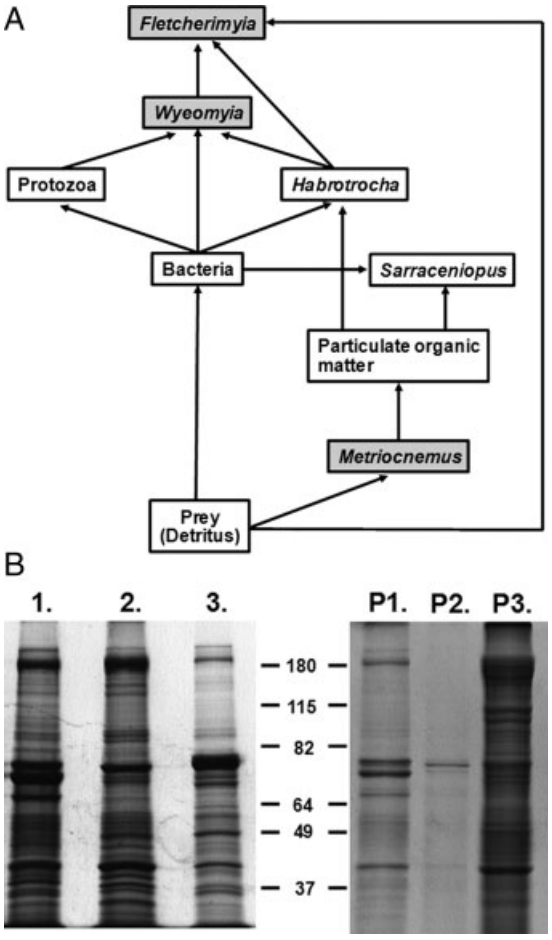

Figure 1. (A) The *Sarracenia* food web is a detritus-based ''brown web'' with a resource input of captured insect prey. The three common dipteran species chosen for proteomic characterization in this study are indicated by shaded boxes. (B-left panel) Extracted proteins from *F. fletcheri* (left lane), *W. smithii* (middle lane), and *M. knabi* (right lane) subjected to SDS-PAGE and stained with coomassie. Each lane was divided into 17 regions and subjected to in-gel tryptic digestion prior to LC-MS/MS analysis. (B-right panel) Extracted proteins for three separate pitcher plant mixed particulates were subjected to SDS-PAGE and stained with coomassie.

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

melanogaster database (Supporting Information Tables 1–3). As expected, many proteins were common to all three species (Supporting Information Fig. 4 and Supporting Information Table 7) because a search of the *D. melanogaster* database would favor the identification of highly conserved proteins. However, a few proteins were uniquely identified in each of the three species (Table 1 and Supporting Information Table 7). Noteworthy are the 71 peptides identifying isoforms of larval serum proteins that were found only in *F. fletcheri*. These peptides were identified primarily from the major coomassie-stained *F. fletcheri* band running at ~78 kDa. Given its predominance, this protein family could potentially be used as a diagnostic indicator protein of *F. fletcheri*. Additionally, a number of proteins were identified in one species by more than threefold the number of peptides found in either one of the other two (Table 1), suggesting that additional proteins, or their increased relative abundance, could also be used as species-specific indicators. It is anticipated that once protein databases are established for these organisms many additional proteins will be identified from the data collected in this study.

Because of the high overlap in identified proteins among species, most of the identified proteins would not serve as species-specific indicators. However, we asked whether the same protein identified from all three species might nonetheless harbor species-specific peptide indicators. If sufficient numbers of such peptides were identified from the relatively invariant residents and prey constituents of the *S. purpurea* food web, these peptides conceivably could serve to trace species and specific proteins even from a mixed protein sample. Such methods would be particularly useful if coupled with quantitative proteomic approaches such as AQUA analyses [11]. It is important to note, however, that the relative abundance of a given protein may only be a rough indicator of the relative species abundance as some proteins are highly regulated at the level of the individual organism.

To search for such indicator peptides, we took the acquired data for the most identified protein in the entire data set, muscle myosin heavy chain. Mass spectral data came from the peptides derived from the dominant band at approximately 220 kDa and its flanking regions. These spectral data were then subjected to a SEQUEST search against a composite database of the most readily identified *D. melanogaster* myosin heavy chain protein sequence concatenated to the next 99 most similar protein sequences identified by BLAST searching. Concatenated to the end of this database were the reversed sequences of all 100 proteins. This entire database is provided as Supporting Information Database 1. The search was conducted as described in the Supporting Information Materials and methods, except trypticity was not

Table 1. Major proteins and signature proteins identified from arthropod associates of *S. purpurea*

Protein name/database reference	*W. smithii*	*M. knabi*	*F. fletcheri*	Total
Myosin heavy chain/Mhc-PG_FBgn0086783	379	469	362	1210
Actin 5C/Act5C-PB_FBgn0000042	74	120	61	255
Tropomyosin 1/Tm1-PK_FBgn0003721	35	56	27	118
α-Actinin/Actn-PC_FBgn0000667	24	39	39	102
ATP synthase-β/ATPsyn-β-PA_FBgn0010217	42	42	7	91
βTub56D/βTub56D-PB_FBgn0003887	35	29	26	90
αTub84D/αTub84B-PA_FBgn0003884	28	37	21	86
SERCA (xalcium ATPase)/Ca-P60A-PE_FBgn0004551	15	54	0	69
Upheld (troponin T)/up-PA_FBgn0004169	16	27	13	56
Tropomyosin 2/Tm2-PA_FBgn0004117	12	22	16	50
Glycogen phosphorylase/GlyP-PA_FBgn0004507	16	13	20	49
Larval serum protein 1 γ/Lsp1γ-PA_FBgn0002564	0	0	39	39
Arginine kinase/Argk-PA_FBgn0000116	18	10	9	37
Elongation factor 1α48D/Ef1α48D-PA_FBgn0000556	10	9	17	36
Transitional endoplasmic reticulum ATPase TER94/TER94-PA_FBgn0261014	25	10	0	35
Heat shock protein 83/Hsp83-PA_FBgn0001233	12	10	13	35
Actin 57B/Act57B-PA_FBgn0000044	4	16	12	32
Larval serum protein 1 α/Lsp1α-PA_FBgn0002562	0	0	32	32
α-Spectrin/α-Spec-PA_FBgn0250789	12	10	5	27
Glutamate dehydrogenase/Gdh-PF_FBgn0001098	3	0	17	20
Haloacid dehalogenase-like hydrolase/CG5177-PA_FBgn0031908	0	0	16	16
Eukaryotic initiation factor-4A/eIF-4a-PA_FBgn0001942	3	3	9	15
Transferrin 1/Tsf1-PA_FBgn0022355	0	0	15	15
Cheerio (Filamin)/cher-PB_FBgn0014141	2	2	10	14
Vacuolar H[+]-ATPase 55kD B subunit, isoform B/Vha55-PB_FBgn0005671	2	9	3	14

Proteins identified by more than 25 peptides for all three species collectively are indicated. Six additional proteins are also listed that were identified collectively by more than ten peptides, and show one species with a threefold higher number of identified peptides than any other for that protein. Numbers indicate the number of peptides identified for each protein in the different sample types. Shaded cells suggest the indicated proteins may be species-dependent signature proteins (more than a 2.5-fold increase observed in one species over any other species). For complete listings, see Supporting Information Tables 1–3.

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

required until after the search was complete for increased stringency. Tryptic peptides from this search were filtered to a <0.1% false discovery rate using stringent Xcorr, ΔCorr, and ppm values [12]. Peptides were then compiled and sorted to identify species-specific peptides (Table 2 and Supporting Information Table 4). A number of sets of homologous muscle myosin heavy chain peptides were identified (Supporting Information Tables 5 and 6) where a peptide from one species was unique, typically by a difference in only one amino acid (Table 2). Furthermore, definitive mass

Table 2. Myosin heavy chain signature peptides for major arthropod associates of *S. purpurea*.

F. fletcheri	*M. knabi*	*W. smithii*
R.QLEEAESQVSQLSK.**I**	R.QLEEAESQVSQLSK.**I**	R.QLE**D**AESQVSQLSK.**I**
K.DIQTALEEEQR.A*	K.DIQTALEEEQR.A*	K.D**V**Q**S**ALEEEQR.A*
K.GGGFATVSSAYK.E	K.GGGFATVSSSYK.E	K.G**A**GFATVSSS**Y**K.E
K.T**G**EELQAAEDK**I**NHLNK.V*	K.TAEELQAAEDKVNHLNK.V*	K.TAEELQAAEDKVNHLNK.V*
K.ELEQTIQR.K	K.ELEQTIQR.K	K.ELEQT**M**R.K
R.KALEQQIK.E*	R.KALE**T**QIK.D*	R.KALEQQIK.E*
K.EL**S**FQSEEDR.K	K.ELTFQSEEDR.K	K.ELTFQSEEDR.K
K.KMQGETNQK.T	K.KMQGETNQK.T	K.KMQGE**V**NQK.T
R.LDEAEANALK.G	R.LDEAEANALK.G	R.LDEAE**S**NALK.G
R.NQLFQ**Q**K.K	R.NQLFQTK.K	R.NQLFQTK.K
R.VISQQ**SL**ER.S	R.VISQQTLER.S	R.VISQQTLER.S
K.LTQEAV**S**DLER.N	K.LTQEAVADLER.N	K.LTQEAVADLER.N
K.HNDAVAEMAEQV**E**QLNK.L	K.HNDAVAEMAEQVDQLNK.L	K.HNDAVAEMAEQVDQLNK.L
R.DLEEANIQHE**S**TLANLR.K	R.DLEEANIQHEGTLANLR.K	R.DLEEANIQHEGTLANLR.K
K.DKELSSITAK.L	K.DKELSSITAK.L	K.DKEI**S**AL**S**AK.L
K.RLADEEAR.E	K.RLADEEAR.E	K.RLADEE**S**R.E
R.ELENELDGEQR.R	R.ELENELDGEQR.R	R.ELE**S**ELD**S**EQR.R
K.DLVGQVNPPK.Y	K.DLVGQVNPPK.Y	K.D**QLQ**QVNPPK.Y
R.SYHIFYQIMSGSV**PGVK**.D	R.SYHIFYQ**M**MSGSV**PGLK**.D	R.SYHIFYQ**M**MSGSV**K**.G
R.ITAAVMHMGGMK.F	R.ITAAVMHMGGMK.F	R.ITAAVMHMGGMK.F
K.LKVDDLAAELDASQK.E	K.LKVDDLAAELDASQK.E	K.LKVDDLAAELDASQK.E
R.LAEAEETIESLNQK.C	R.LAEAEETIESLNQK.C	R.LAEAEETIESLNQK.C
R.IEELEEEVEAER.Q	R.IEELEEEVEAER.Q	R.IEELEEEVEAER.Q
K.TALLDSLSGEK.G	K.TALLDSLSGEK.G	K.TALLDSLSGEK.G
R.IQEKEEEFENTR.K	R.IQEKEEEFENTR.K	R.IQEKEEEFENTR.K
R.LEEAGGATSAQIELNK.K	R.LEEAGGATSAQIELNK.K	R.LEEAGGATSAQIELNK.K
R.LSTEVEDLQLEVDR.A	R.LSTEVEDLQLEVDR.A	R.LSTEVEDLQLEVDR.A
K.NLADEVK.D	K.NLADEVK.D	K.NLADEVK.D
R.ANALQNELEESR.T	R.ANALQNELEESR.T	R.ANALQNELEESR.T
R.AQLELSQVR.Q	R.AQLELSQVR.Q	R.AQLELSQVR.Q
R.MQDLVDK.L	R.MQDLVDK.L	R.MQDLVDK.L
R.NIHEIEK.A	R.NIHEIEK.A	R.NIHEIEK.A
K.KLEADINELEIALDHANK.A	K.KLEADINELEIALDHANK.A	K.KLEADINELEIALDHANK.A
K.KLSIENSDLLR.Q	K.KLSIENSDLLR.Q	K.KLSIENSDLLR.Q
R.NLEHDLDNLR.E	R.NLEHDLDNLR.E	R.NLEHDLDNLR.E
R.NYSTELFR.L	R.NYSTELFR.L	R.NYSTELFR.L
R.QIEEAEEIAALNLAK.F	R.QIEEAEEIAALNLAK.F	R.QIEEAEEIAALNLAK.F
R.AGVLGQMEEFRDER.L	R.AGVLGQMEEFRDER.L	R.AGVLGQMEEFRDER.L
R.ALDSMQASLEAEAK.G	R.ALDSMQASLEAEAK.G	R.ALDSMQASLEAEAK.G
K.GSLEDQVVQTNPVLEAFGNAK.T	K.GSLEDQVVQTNPVLEAFGNAK.T	K.GSLEDQVVQTNPVLEAFGNAK.T
R.AEQDHAQTQEK.L	R.AEQDHAQTQEK.L	R.AEQDHAQTQEK.L
K.ANAEAQLWR.Q	K.ANAEAQLWR.Q	K.ANAEAQLWR.Q
K.AKLEQTLDELEDSLER.E	K.AKLEQTLDELEDSLER.E	K.AKLEQTLDELEDSLER.E
K.RYQQQLK.D	K.RYQQQLK.D	K.RYQQQLK.D
R.TLLEQADR.G	R.TLLEQADR.G	R.TLLEQADR.G
K.DLLDQIGEGGR.N	K.DLLDQIGEGGR.N	K.DLLDQIGEGGR.N
K.GAYEEGQEQLEAVR.R	K.GAYEEGQEQLEAVR.R	K.GAYEEGQEQLEAVR.R

MS/MS data from the major bands harboring myosin heavy chain (MHC) as determined by searching the *D. melanogaster* database were then searched against an assembled database containing 100 metazoan MHC orthologs (see text for details). A total of 47 sets of homologous peptides were identified in all three species. Shaded cells indicate the 19 peptides that show species-specific signatures. Distinguishing amino acids are in bold. Asterisks and underlining denote peptide sequences for which MS/MS spectra are provided in Supporting Information Figs. 1–3. A complete listing of all identified peptides from the MHC search are listed in Supporting Information Table 5

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

spectra were obtained for these unique peptides (examples are shown in Supporting Information Figs. 1–3). However, there remains a formal possibility that these peptides are not species-specific for the following reasons. First, although data for each of the arthropod associates were from a mixture of several individual isolates mixed together, the MS data were generated from a single gel (Fig. 1B) and as such will need to be further established in future proteomic analyses. Second, if unlike Drosophila, these insects harbor two or more muscle myosin heavy chain genes, then we cannot exclude the possibility that the sequence of a second muscle myosin protein, perhaps in low abundance in our samples, negates a perceived uniqueness. However, our data do not support this as individual arthropods did not show multiple homologous peptides covering the same regions of myosin heavy chain. Third, although these peptides may distinguish between the three insects examined here, other insects not examined may contribute to the heterogeneous biomass of the pitcher plant community and these insects may have protein sequences that also negate uniqueness with the pitcher plant ecosystem. This last possibility is likely to be true to a degree. However, if multiple peptides are monitored, stoichiometric measurements could be used to identify sets of peptides belonging uniquely to one organism. Furthermore, for a state analysis of a dynamic ecosystem it is not a requirement to know the species origin of a given peptide biomarker. It would be sufficient to know the relative abundance of biomarker peptides if they provided predictive power regarding the ecosystem state and dynamics. Ultimately, we plan to identify peptides that serve these goals and our initial characterization here suggests that significant identification and precision can be achieved.

In conclusion, in spite of the lack of database representation for these three dipterans, our analyses identified apparent species-specific peptide biomarkers from a database of related proteins. Our approach should be applicable toward the identification of numerous species-specific peptide biomarkers in this specific food web as well as in other systems for which protein databases do not yet exist. It is important to note that while simple species identification can be achieved readily by inexpensive methods such as light microscopy, an analysis of the individual components of the ecosystem biomass will provide molecular information that may be more acutely regulated than whole organisms and may be more sensitive indicators of ecosystem state. Once specific peptide biomarkers are confirmed, quantitative MS can facilitate their absolute quantification and determine how these proteins or peptides change in abundance when the ecosystem is in different states. Furthermore, specific markers may facilitate the evaluation of the protein constituent of ecosystem services, which is still an unexplored area of study.

All MS data from this study are freely available in the PRIDE database (http://www.ebi.ac.uk/pride) under the accession numbers 16271–16273 and can be viewed using the freely available PRIDE Inspector program (http://code.google.com/p/pride-toolsuite/wiki/PRIDEInspector).

This work was supported by a University of Vermont Innovation Research Grant (N.J.G. and B.A.B.); the Vermont Genetics Network through NIH Grant Number P20 RR16462 from the INBRE Program of the NCRR (B.A.B.); and by NSF grants DEB-0541936 and DEB-0546180 (N.J.G. and A.M.E.).

The authors have declared no conflict of interest.

References

[1] Costanza, R., d'Arge, R., deGroot, R., Farber, S. et al., The value of the world's ecosystem services and natural capital. *Nature* 1997, *387*, 253–260.

[2] Kremen, C., Managing ecosystem services: what do we need to know about their ecology? *Ecol. Lett.* 2005, *8*, 468–479.

[3] Kammenga, J. E., Herman, M. A., Ouborg, N. J., Johnson, L., Breitling, R., Microarray challenges in ecology. *Trend Ecol. Evol.* 2007, *22*, 273–279.

[4] Moe, S. J., Stelzer, R. S., Forman, M. R., Harpole, W. S. et al., Recent advances in ecological stoichiometry: insights for population and community ecology. *Oikos* 2005, *109*, 29–39.

[5] Ellison, A. M., Gotelli, N. J., Brewer, J. S., Cochran-Stafira, D. L. et al., The evolutionary ecology of carnivorous plants. *Adv. Ecol. Res.* 2003, *33*, 1–74.

[6] Ellison, A. M., Gotelli, N. J., Energetics and the evolution of carnivorous plants – Darwin's 'most wonderful plants in the world'. *J. Exp. Bot.* 2009, *60*, 19–42.

[7] Addicott, J. F., Predation and prey community structure: an experimental study of the effect of mosquito larvae on the protozoan community of pitcher plants. *Ecology* 1974, *55*, 475–492.

[8] Bradshaw, W. E., Creelman, R. A., Mutualism between the carnivorous purple plant *Sarracenia purpurea* and its inhabitants. *Am. Midland Nat.* 1984, *112*, 294–304.

[9] Buckley, H. L., Miller, T. E., Ellison, A. M., Gotelli, N. J., Reverse latitudinal trends in species richness of pitcher-plant food webs. *Ecol. Lett.* 2003, *6*, 825–829.

[10] Ballif, B. A., Cao, Z. W., Schwartz, D., Carraway, K. L., Gygi, S. P., Identification of 14-3-3 epsilon substrates from embryonic murine brain. *J. Proteome Res.* 2006, *5*, 2372–2379.

[11] Gerber, S. A., Kettenbach, A. N., Rush, J., Gygi, S. P., The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. *Methods Mol. Biol.* 2007, *359*, 71–86.

[12] Elias, J. E., Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 2007, *4*, 207–214.

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim