

# Genome-Wide Analysis of Adaptive Molecular Evolution in the Carnivorous Plant *Utricularia gibba*

Lorenzo Carretero-Paulet<sup>1,\*</sup>,†, Tien-Hao Chang<sup>1,†</sup>, Pablo Librado<sup>2</sup>, Enrique Ibarra-Laclette<sup>3,4</sup>, Luis Herrera-Estrella<sup>3</sup>, Julio Rozas<sup>2</sup>, and Victor A. Albert<sup>1,\*</sup>

<sup>1</sup>Department of Biological Sciences, University at Buffalo

<sup>2</sup>Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Spain

<sup>3</sup>Laboratorio Nacional de Genómica para la Biodiversidad-Langebio/Unidad de Genómica Avanzada UGA, Centro de Investigación y Estudios Avanzados del IPN, Irapuato, Guanajuato, México

<sup>4</sup>Present address: Red de Estudios Moleculares Avanzados, Instituto de Ecología A.C., Xalapa, Veracruz, México

\*Corresponding author: E-mail: lcpaulet@gmail.com; vaalbert@buffalo.edu.

†These authors contributed equally to this work.

Accepted: December 19, 2014

## Abstract

The genome of the bladderwort *Utricularia gibba* provides an unparalleled opportunity to uncover the adaptive landscape of an aquatic carnivorous plant with unique phenotypic features such as absence of roots, development of water-filled suction bladders, and a highly ramified branching pattern. Despite its tiny size, the *U. gibba* genome accommodates approximately as many genes as other plant genomes. To examine the relationship between the compactness of its genome and gene turnover, we compared the *U. gibba* genome with that of four other eudicot species, defining a total of 17,324 gene families (orthogroups). These families were further classified as either 1) lineage-specific expanded/contracted or 2) stable in size. The *U. gibba*-expanded families are generically related to three main phenotypic features: 1) trap physiology, 2) key plant morphogenetic/developmental pathways, and 3) response to environmental stimuli, including adaptations to life in aquatic environments. Further scans for signatures of protein functional specialization permitted identification of seven candidate genes with amino acid changes putatively fixed by positive Darwinian selection in the *U. gibba* lineage. The *Arabidopsis* orthologs of these genes (*AXR*, *UMAMIT41*, *IGS*, *TAR2*, *SOL1*, *DEG9*, and *DEG10*) are involved in diverse plant biological functions potentially relevant for *U. gibba* phenotypic diversification, including 1) auxin metabolism and signal transduction, 2) flowering induction and floral meristem transition, 3) root development, and 4) peptidases. Taken together, our results suggest numerous candidate genes and gene families as interesting targets for further experimental confirmation of their functional and adaptive roles in the *U. gibba*'s unique lifestyle and highly specialized body plan.

**Key words:** genome evolution, phenotypic diversification, carnivorous plants, gene family expansions and contractions, positive selection.

## Introduction

The carnivorous syndrome has evolved multiple times among the flowering plants as a specialized solution for nutrient acquisition in phosphorus- and nitrogen-limited environments (Ellison and Gotelli 2009). Although the fascinating trapping strategies and specialized body designs of carnivorous plants have attracted scientific interest for centuries (Darwin 1875), the underlying genetic basis of the carnivorous syndrome remains largely unknown. The asterid family Lentibulariaceae is the largest and most phenotypically diverse carnivorous plant family, comprising three genera with distinct trapping

mechanisms: *Pinguicula*, *Genlisea*, and *Utricularia* (Juniper et al. 1989; Taylor 1989). *Utricularia*, the bladderworts, are the largest and most diverse of all carnivorous plants. *Utricularia* species typically live in nutrient-poor aquatic, terrestrial, or epiphytic environments (Chao 2003), where they supplement normal photolithotrophic nutrition by trapping various microscopic prey animals (Givnish 1989). The trapping structures of *Utricularia*, the bladder-like suction traps, rank among the most complex leaf structures known in the plant kingdom (Juniper et al. 1989). These bladders can appear on apparently dissimilar parts of the plant body in different

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

species, and their morphologies are diverse. Other than concerning the traps, a single *Utricularia* body plan is difficult to define; the boundaries between classical organ identities are blurred to the extent that leaf versus shoot becomes a difficult diagnosis for many species, especially given their lack of true roots (Adlassnig et al. 2005).

The recent publication of the nuclear genome of *Utricularia gibba*, an obligate aquatic bladderwort with thread-like and highly ramified vegetative organs, has revealed several interesting genomic architectural features (Ibarra-Laclette et al. 2013). Interestingly, the tiny *U. gibba* genome (82 Mb) results from the shrinkage of noncoding DNA, and not from an appreciable reduction in overall gene number (Ibarra-Laclette et al. 2013). Despite this comparable gene number, *U. gibba* has undergone three whole-genome duplication (WGD) events since its common ancestry with grape (Ibarra-Laclette et al. 2013). As a result, vast genetic resources were introduced to its genome, leading to multiple opportunities for possible functional specialization of gene duplicates. Therefore, it is expected that gene turnover has had an impact on *U. gibba*'s unique lifestyle, especially when its present tiny genome size suggests a strong constraint on its genetic content.

The evolution and architecture of plant genomes is greatly shaped by recurrent events of gene duplication and loss events (Lynch and Conery 2000; Blanc and Wolfe 2004; Cui et al. 2006; Fei et al. 2006; Jiao et al. 2011; Proost et al. 2011), which are believed to play a key role in generating plant phenotypic diversity and speciation, including the explosive early diversification of angiosperms (De Bodt et al. 2005; Soltis and Burleigh 2009; Amborella Genome Project 2013). Gene duplication provides new substrate for mutation and selection to act upon. In most cases, a new gene duplicate evolves neutrally, stochastically accumulating loss of function mutations (pseudogenization). However, a fraction of duplicates might be retained when 1) gene duplication endows organisms with mutational robustness as a result of functional redundancy (Gu et al. 2003), 2) selection for increased gene dosage occurs (Conant and Wolfe 2008), or 3) accompanying acquisition of novel or specialized functions. From a population genetic perspective, relaxed purifying selection and positive Darwinian selection may contribute to functional specialization of gene duplicates (Moore and Purugganan 2003; Zhang 2003; Conant and Wolfe 2008). Positive selection (PS) promotes the fixation of mutations with advantageous fitness effects, while relaxed purifying selection can tolerate the fixation of specific mutations that may also contribute to the acquisition of new or specialized functions.

In this study, we classified genes from the *U. gibba*, *Mimulus guttatus* (*Mimulus*), *Solanum lycopersicum* (tomato), *Arabidopsis thaliana* (*Arabidopsis*), and *Vitis vinifera* (grape) genomes by clustering orthologous and in-paralogous genes from these five species into orthogroups, which we take to be equivalent to gene families. Gene families were

classified as expanded/contracted in specific lineages or stable in size by means of the maximum likelihood (ML) framework provided by Badirate (Librado et al. 2012). We found that the orthogroups significantly expanded in the *U. gibba* lineage may have adaptive significance for the evolution of the species' unique phenotypic features. Moreover, to identify signatures of adaptive functional specialization at the protein-coding sequence level, we performed a preliminary genome-wide scan for PS on the data set of co-orthologous groups that included representatives from the five species, and examined in further detail seven of the genes showing signatures of PS in the *U. gibba* lineage. Our results provide key genes and gene families that may be interesting targets for further experimental confirmation of their functional and adaptive roles in *U. gibba*'s unique lifestyle and highly specialized body plan.

## Materials and Methods

### Functional Annotation of Five Plant Eudicot Genomes and Statistical Analysis of Gene Ontology Term Enrichment/Depletion

All gene models predicted in the fully sequenced genomes of *Arabidopsis*, grape, *U. gibba*, *Mimulus*, and tomato were downloaded from CoGe (<http://genomeevolution.org/CoGe/>) and functionally annotated by assigning their associated generic gene ontology (GO) terms and enzyme codes (EC) through the BLAST2GO program (Conesa et al. 2005), and integrating information about the occurrence of INTERPRO functional domains identified by INTERPROSCAN (Zdobnov and Apweiler 2001) and KEGG EC biochemical pathways (Ogata et al. 1999). Annotations were further expanded using ANNEX (Myhre et al. 2006). The following settings were used: BLAST searches were conducted for each protein (BLASTX, nr database, HSP cutoff length 33, report 20 hits, maximum *E*-value 1 E-10), followed by mapping and annotation (*E*-value hit filter 1 E-10, annotation cutoff 55, GO weight 5, HSP-hit coverage cutoff 20). As a gene might have more than one function, it might be annotated with more than one GO functional category or EC code. A substantial fraction of genes per species was associated with at least one GO generic term, ranging from 60.61% to 82.30% for *U. gibba* and *Arabidopsis* genes, respectively, averaging 4.62–5.44 associated GO terms per functionally annotated gene (table 1). To obtain a broader overview of functional annotation, the resulting generic GO terms were mapped onto the corresponding Plant GO slim terms. The GO and GOslim annotations of the five genomes are available as [supplementary files S1 and S2, Supplementary Material](#) online, respectively. We performed significance analyses of GO term enrichment/depletion of subsets of genes versus all genes in a genome by means of Fisher's exact tests. To control for multiple testing,

**Table 1**

Numbers of Genes and GO Functional Annotations for Five Eudicot Plant Genomes

	<i>Utricularia gibba</i>	<i>Arabidopsis</i>	Grape	<i>Mimulus</i>	Tomato
Total number of genes	28,494	27,204	26,346	27,501	34,727
Total number of annotated genes	17,270	22,390	19,090	21,435	22,651
Annotated genes (%)	60.61	82.30	72.46	77.94	65.23
Total number of generic GO terms	79,933	118,023	92,018	102,601	107,423
Average number of generic GO terms/gene	4.63	5.27	4.82	4.79	4.74
Total number of plant GO slim terms	83,160	121,977	95,293	106,069	111,193
Average number of plant GO slim terms/gene	4.82	5.45	4.99	4.95	4.91

the resulting  $P$ -values were corrected according to Benjamini and Hochberg (1995).

### Gene Family Definition Using OrthoMCL

The complete proteomes encoded by the genomes of *Arabidopsis*, grape, *U. gibba*, *Mimulus*, and tomato were globally compared. The complete data set represents 144,274 protein-coding sequences. An all-against-all comparison was performed using BLASTP ( $1 \times 10^{-10}$ ) followed by clustering with OrthoMCL v1.4 (Li et al. 2003) with default parameters. To avoid possible enrichment of transposable element in our analyses of lineage-specific expansions, we filtered out orthogroups containing sequences with significant similarity to them ( $E$ -values  $< 10^{-5}$ , bit score  $\geq 45$  in BLASTX searches against RepBase v16.03 database [Jurka et al. 2005]). The final data set (supplementary file S3, Supplementary Material online) contained 17,324 multigene orthogroups and 93,865 proteins.

### Lineage-Specific Gene Family Size Changes

To analyze the evolution of gene families, we applied gain and death (GD) stochastic models as implemented in the BadiRate program (Librado et al. 2012). BadiRate provides the appropriate statistical framework for testing biologically relevant hypothesis on the input data, consisting of the species phylogenetic tree and the sizes of gene families in each extant species. To detect species-specific rates of gene GD, we compared the fit of different branch models to the data applying the same approach as in Denoeud et al. (2014).

### Functional/Selective Constraints in Protein-Coding Regions

Selection may differentially act on nonsynonymous (amino acid-changing;  $d_N$ ) substitutions compared with the putatively neutral synonymous (silent;  $d_S$ ) ones. Under the neutral model of evolution, synonymous and nonsynonymous substitutions accumulate at the same rate ( $\omega = d_N/d_S \sim 1$ ; Ohno 1970; Li et al. 1981; Nei 2005). Conversely, deviations from this pattern indicate the action of purifying selection (which purges deleterious mutations to conserve the protein structure and function;  $\omega \ll 1$ ), or the action of PS promoting the fixation of

nonsynonymous mutations with advantageous fitness effects ( $\omega > 1$ ). Estimation of  $\omega$  therefore yields insights on the molecular evolutionary mechanisms of protein diversification and functional specialization.

We estimated  $\omega$  values by means of the codeml program in the PAML v4.4 package (Yang 1997) on the basis of multiple alignments of orthologous codon sequences and a tree topology. To obtain final codon alignments, the coding sequences were aligned with PRANK (Löytynoja and Goldman 2005) on the GUIDANCE server (<http://guidance.tau.ac.il/>; Penn et al. 2010), and the default cutoff of 0.93 confident score was applied for the removal of unreliably aligned positions.

Two different classes of models were implemented in the codeml program: “branch-specific” (which permit a different  $\omega$  ratio in *U. gibba*; Yang 1998) and “branch-site” models (which also accounting for uneven selective pressures along sequences; Yang and Nielsen 2002). By comparing the fit to the data of two nested models via a likelihood ratio test, we tested for: 1) asymmetric sequence evolution (one-ratio model 0 versus the branch-specific model described above); 2) heterogeneous selective pressures across sites ( $\omega$  ratio constant among branches but variable among sites versus the clade model described in Bielawski and Yang (2004)); and 3) PS ( $\omega$  fixed at 1 at the *U. gibba* branch versus a model featuring an extra class of sites with  $\omega > 1$ ; Yang et al. 2005).

In a first approach, we used the original, entire set of 8,676 shared orthogroups containing homologs from all five species. Prior to the analyses, we discarded 1) sequences whose alignments spanned less than 150 nt or showed an identity score  $< 30\%$  (Li et al. 2001) (1,659 orthogroups), as too divergent and, 2) sequences containing potential errors leading to in-frame codon stops (169 orthogroups). This resulted in a set of 6,848 shared orthogroups well conserved among all five species. The entire process was automated using a PERL pipeline adapted from Carretero-Paulet and Fares (2012).

We further confirmed a subset of seven selected *U. gibba* genes detected as under PS in our previous analysis. For this purpose, we performed additional analyses of PS by sampling orthologs from a greater number of species (supplementary table S7, Supplementary Material online) and their corresponding tree topologies as returned from the NCBI taxonomy tree tool (<http://www.ncbi.nlm.nih.gov/Taxonomy/Common>

Tree/wwwcmt.cgi). The orthology relationships of the sequences sampled were first assessed using COGE/GEvo (<http://genomeevolution.org/CoGe/GEvo.pl>) to ascertain strong evidence for synteny, which was considered confirmation of (co-)orthology. In some cases, codon sequences were repredicted from the corresponding genomic sequences through comparisons with their orthologous proteins using GENEWISE (Birney and Durbin 2000).

### 3D Protein Structure Modeling

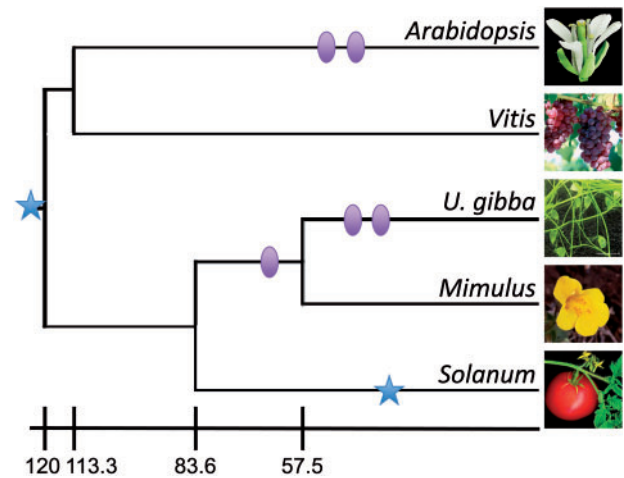
Protein structure models were generated using the SWISS-MODEL workspace (Arnold et al. 2006) on alignment mode and specific structure template selection. The structures of the *U. gibba* orthologs to AXR1, IGS, TAR2, SOL1, DEG9, and DEG10 were modeled using the protein structures with PDB identifiers 2nvuA (2.80 Å) (Huang et al. 2007), 3tsmA (2.15 Å), 3bwnB (2.25 Å) (Tao et al. 2008), 1h8IA (2.60 Å) (Aloy et al. 2001), and 4flnB (2.80 Å) (Sun et al. 2012), respectively, as templates. Protein 3D-structural models were displayed and edited using VMD 1.9.1 (Humphrey et al. 1996).

## Results and Discussion

### Analysis of Gene Family Turnover in *U. gibba*

Genes in the genomes of *U. gibba*, the asterids *Mimulus* and tomato, and the rosids *Arabidopsis* and grape, which represent highly diverse core eudicot lineages with different histories of WGD (fig. 1 and table 1) were classified into groups of orthologous genes (orthogroups) using OrthoMCL. We used the resulting set of 17,324 orthogroups to examine the relationship between the *U. gibba* genome's compactness and its rates of gene turnover, while accounting for their potential adaptive value for the *U. gibba* evolution. We evaluated the fits of different branch models of gene turnover to each orthogroup using the statistical framework provided by BadiRate (Librado et al. 2012). According to the ML best-fit branch model of gene turnover, each orthogroup was classified as 1) lineage-specific expanded or contracted, or 2) stable in size across species (supplementary table S1, Supplementary Material online). Six hundred twenty-six orthogroups (comprising 1,754 *U. gibba* genes) were considered significantly expanded in the *U. gibba* lineage (supplementary tables S2 and S3, Supplementary Material online), whereas 628 orthogroups (comprising only eight *U. gibba* genes) were classified as contracted, having in most cases lost all of their *U. gibba* orthologs (supplementary tables S4 and S5, Supplementary Material online). Six thousand one hundred seventy orthogroups (comprising 6,188 *U. gibba* genes) corresponded to orthogroups showing constant sizes across all five genomes.

Although some genes could be incorrectly annotated or biologically not meaningful, GO annotations represent the best approach developed so far to gain further insights into the potential adaptive value of the *U. gibba*-specific



**Fig. 1.**—Phylogenetic tree depicting the taxonomic relationships among the five plant species with fully sequenced genomes examined in this study. Branch lengths reflect evolutionary time (in millions of years). For the timing of these events, we used estimates from TimeTree (Hedges et al. 2006), which are shown next to the scale bar. The history of WGDs is mapped onto the tree, with circles and stars representing WGDs and duplications, respectively. The positions of these events are not meant to reflect their timing of occurrence.

expansions. In particular, we performed a GO enrichment analysis (GO terms in genes belonging to expanded orthogroups versus all GO terms in the *U. gibba* genome), partitioned by either generic GO terms or Plant GO slim terms (table 1 and supplementary tables S2 and S3, Supplementary Material online). Seven GO slim terms (out of a total of 104, including two significant after correction for multiple testing) and 94 generic GO terms (out of a total of 4,330, including 21 significant after correction for multiple testing) were found to be significantly and specifically enriched among *U. gibba*-expanded orthogroups (table 2). Given the compact genome size of *U. gibba*, the expansion of specific gene families may reflect the emergence of key *U. gibba* adaptations, such as physiological and morphological phenotypic plasticity (assuming that our GO annotations are essentially correct, and some degree of uncertainty in specific GO terms with a low number of genes). After careful examination of the GO terms found as enriched in expanded orthogroups, we operationally grouped them into three main biological functions that are explored in more detail below. These biological functions are: 1) trap physiology, 2) key plant morphogenetic/developmental pathways, and 3) response to environmental stimuli, including adaptations to life in aquatic environments.

### The Putative Role of Gene Family Expansion in the Evolution of Trap Physiology in *U. gibba*

To understand the molecular evolutionary basis of the *U. gibba* carnivorous syndrome, we explored functional

Table 2

Generic and Plant GO Slim Terms Enriched among Gene Families Expanded in *Utricularia gibba*

GO	GO Description	SC	GC	P-Value
<i>Generic GO term</i>				
B	Cuticle development	11	22	1.30E-08
B	Very long-chain fatty acid metabolic process	11	22	1.30E-08
B	Abaxial cell fate specification	8	13	3.93E-07
B	Vegetative phase change	8	16	1.29E-06
B	Response to chitin	15	78	1.30E-06
B	Regulation of transcription, DNA-templated	31	298	1.71E-06
M	Chlorophyll binding	10	36	4.76E-06
B	Photosynthesis, light harvesting	9	28	5.28E-06
B	Cell wall biogenesis	8	23	1.10E-05
M	Fatty acid elongase activity	5	6	2.39E-05
B	Response to oxidative stress	18	145	2.97E-05
B	Rhamnogalacturonan II biosynthetic process	5	7	3.97E-05
C	Fatty acid elongase complex	5	8	6.25E-05
B	Response to light stimulus	13	87	6.86E-05
M	UDP-xylosyltransferase activity	5	9	9.44E-05
B	Unidimensional cell growth	11	71	0.000182731
B	Sphingoid biosynthetic process	4	6	0.000301803
M	Sphingosine hydroxylase activity	4	6	0.000301803
B	Protein-chromophore linkage	7	30	0.000321316
C	Membrane	65	1067	0.000407321
B	Transport	18	190	0.000636728
C	Extrachromosomal circular DNA	3	3	0.000871807
B	Regionalization	3	3	0.000871807
B	Plant-type cell wall biogenesis	5	17	0.000974114
B	Pollen tube development	5	18	0.001207735
M	Cyclin-dependent protein kinase inhibitor activity	4	10	0.001280476
C	Ubiquitin ligase complex	8	52	0.00140315
C	Photosystem II antenna complex	3	4	0.001484478
M	Cellulase activity	4	12	0.002196988
C	Cytochrome b6f complex	3	5	0.00231117
C	PSII associated light-harvesting complex II	3	5	0.00231117
M	Pectate lyase activity	5	23	0.003035289
M	Sialyltransferase activity	3	6	0.003373532
B	Organ development	4	15	0.004290413
B	Nonphotochemical quenching	3	7	0.004689987
C	Chloroplast stromal thylakoid	3	7	0.004689987
M	Glycine C-acetyltransferase activity	2	2	0.007475192
M	oxidoreductase activity, acting on CH-OH group of donors	2	2	0.007475192
B	Regulation of gibberellin biosynthetic process	2	2	0.007475192
B	Ammonium transmembrane transport	3	10	0.010305104
C	Calcineurin complex	2	3	0.012159973
B	Purine nucleoside transmembrane transport	2	3	0.012159973
M	Purine nucleoside transmembrane transporter activity	2	3	0.012159973
B	Cytokinin transport	2	3	0.012159973
B	Stomatal lineage progression	2	3	0.012159973
B	Regulation of glycolysis	2	3	0.012159973
B	Chloroplast accumulation movement	2	3	0.012159973
B	Chloroplast avoidance movement	2	3	0.012159973
B	Ammonium transport	3	11	0.012766367
B	Positive regulation of cell proliferation	3	11	0.012766367
B	Cell differentiation	4	23	0.015445481
B	Maintenance of floral meristem identity	2	4	0.017804056
B	Floral whorl development	2	4	0.017804056

(continued)

Table 2 Continued

GO	GO Description	SC	GC	P-Value
B	Adaxial/abaxial axis specification	2	4	0.017804056
C	Piccolo NuA4 histone acetyltransferase complex	2	4	0.017804056
M	C-4 methylsterol oxidase activity	2	4	0.017804056
M	Profilin binding	2	4	0.017804056
B	Hypotonic salinity response	2	4	0.017804056
B	Xenobiotic catabolic process	2	4	0.017804056
M	Translation repressor activity, nucleic acid binding	2	4	0.017804056
C	Cell-cell junction	2	4	0.017804056
M	Threonine synthase activity	2	4	0.017804056
B	Respiratory burst involved in defense response	2	4	0.017804056
B	Response to temperature stimulus	3	13	0.018611667
B	Gravitropism	3	13	0.018611667
M	Chaperone binding	4	25	0.01974656
B	Microtubule-based process	3	14	0.022001797
B	Regulation of protein kinase activity	3	14	0.022001797
C	Peripheral to membrane of membrane fraction	2	5	0.024331887
B	Retinol metabolic process	2	5	0.024331887
B	Protein autoubiquitination	2	5	0.024331887
B	Triglyceride biosynthetic process	2	5	0.024331887
M	Retinal dehydrogenase activity	2	5	0.024331887
B	Protein import into mitochondrial matrix	2	5	0.024331887
B	negative regulation of transcription, DNA-dependent	6	56	0.024501293
B	Lateral root development	3	15	0.025704894
M	Transcription cofactor activity	3	16	0.02972011
M	Double-stranded RNA binding	3	16	0.02972011
C	Integral to plasma membrane	3	16	0.02972011
B	mRNA export from nucleus	2	6	0.031672126
B	Negative regulation of cell proliferation	2	6	0.031672126
M	Adenyl-nucleotide exchange factor activity	2	6	0.031672126
B	Response to blue light	4	30	0.033469207
M	Drug transmembrane transporter activity	5	46	0.036730817
C	NuA4 histone acetyltransferase complex	2	7	0.039757439
M	tRNA dihydrouridine synthase activity	2	7	0.039757439
M	UDP-glucosyltransferase activity	2	7	0.039757439
B	Positive regulation of development, heterochronic	2	7	0.039757439
B	Meristem maintenance	2	7	0.039757439
B	Threonine biosynthetic process	2	7	0.039757439
M	3-deoxy-7-phosphoheptulonate synthase activity	2	7	0.039757439
B	Regulation of translation	2	8	0.048524302
M	Transferase activity, transferring nitrogenous groups	2	8	0.048524302
B	Carpel development	2	8	0.048524302
<i>GO slim term</i>				
C	<i>Intracellular</i>	75	1264	0.000289058
B	<i>Photosynthesis</i>	16	187	0.003647361
B	Multicellular organismal development	41	737	0.019965362
M	Translation regulator activity	2	6	0.031489331
B	Response to abiotic stimulus	46	877	0.032774883
B	Cell differentiation	17	269	0.053689563
C	Intracellular	75	1264	0.000289058

NOTE.—B, M and C indicate biological process, molecular function and cellular component GO classes, respectively. Sample counts and genome counts indicate the total numbers of genes annotated with particular GO terms among expanded families (of which there are 1,754 genes total) versus the entire *U. gibba* genome (which contains a total of 28,494 genes). *P*-values resulting from Fisher exact tests are shown (in italic, those significant after Benjamini–Hochberg correction). Selected GO terms generically related to 1) trap physiology, 2) plant morphogenetic/developmental pathways, and 3) response to environmental stimuli and adaptations to life in aquatic environments are colored orange, green, and blue, respectively.

enrichments among expanded orthogroups that may be related to trap physiology. Since *U. gibba* can acquire nutrients from prey captured by its bladder traps, the digestion of these organisms is a critical function. Various enzymes, such as hydrolases, peptidases, phosphatases, and chitinases have been proposed to be relevant to prey digestion among different carnivorous plant lineages (Sirová et al. 2003; Muller et al. 2004, 2006; Plachno et al. 2006; Hatano and Hamada 2008; Rottloff et al. 2011; Renner and Specht 2013), which independently evolved across angiosperm phylogeny (Ellison and Gotelli 2009).

We found that the generic GO term “response to chitin” was significantly enriched after correction among gene families expanded in the *U. gibba* lineage (table 2 and [supplementary table S2, Supplementary Material](#) online). Chitin is the main component of the cell walls of fungi and also the exoskeleton of arthropods and nematodes, and is a well-known elicitor of plant defense responses (Shibuya and Minami 2001; Kaku et al. 2006). Subclass I chitinases have been proposed to play a major role in pathogenic response in plants (Salzer et al. 2000), and have also been suggested to serve plant carnivory in the order Caryophyllales (Renner and Specht 2012). Although no GO term specifically related to “chitinase” was enriched among the *U. gibba* gene family expansions, the enrichment of the response to chitin GO term may reflect adaptation to chitin detection and activation of relevant biochemical pathways involved in prey digestion, nutrient absorption, and/or in defense against pathogenic fungi.

The production of enzymes for prey digestion and nutrient acquisition, together with the lower photosynthetic capacity of trap biomass compared with leaf-like structures in *Utricularia*, implies a high energetic cost (Adamec 2006). Indeed, a delicate balance between energetic costs and benefits of prey digestion and nutrient acquisition has been proposed for bladderworts (Ellison and Gotelli 2001; Laakkonen et al. 2006). Previous research has suggested that respiratory rates are high in *Utricularia* traps, and that this could result from adaptive protein changes in cytochrome c oxidase (COX), a primary enzyme of the electron transport chain (Laakkonen et al. 2006). It is well known that high charge differences between the mitochondrial lumen and intermembrane space can lead to leakage of electrons and the production of reactive oxygen species (ROS; Turrens 2003; Albert et al. 2010). Interestingly, the generic GO term “response to oxidative stress” was found to be specifically enriched (significant after correction) among the *U. gibba*-expanded gene families (table 2 and [supplementary table S2, Supplementary Material](#) online). The potential adaptive value of this expansion is consistent with previous transcriptomic studies in *U. gibba* that support the hypothesis that COX-derived ROS may account for both increased nucleotide substitution rates and genome-size dynamism following DNA double-strand break repair (Ibarra-Laclette et al. 2011).

We also found the generic GO term “respiratory burst involved in defense response” specifically enriched among *U. gibba*-expanded gene families. This GO term exclusively annotates to ORTHOMCL15238, an orthogroup entirely composed of three *U. gibba* PUB23-like E3 ubiquitin-protein ligases. *PUB23*, similarly to other *PUB* family genes in *Arabidopsis*, likely regulates drought and plant pathogen immunity signaling pathways (Cho et al. 2008). Interestingly, the enhanced respiratory burst in *PUB* mutants might be triggered by chitin (Trujillo et al. 2008). Respiratory burst in response to pathogens can rapidly release ROS such as superoxide and hydrogen peroxide, which play roles in both direct killing of pathogens and induction of defense gene expression (Bolwell 1999). Because plant carnivory may have evolved from defense mechanisms against herbivores and pathogens (Mithofer 2011; Renner and Specht 2013), the possible pathogen defensive role of these *PUB23*-like genes may have been recruited to *U. gibba*'s carnivorous syndrome, perhaps in response to prey capture. The expansion of the *PUB23*-like gene family in *U. gibba* is also consistent with increased ROS levels in traps versus vegetative structures.

#### The Putative Role of Gene Family Expansion in the Evolution of *U. gibba*'s Specialized Body Plan

*Utricularia gibba* shows a unique and highly specialized body plan, including absence of roots and presence of fibrous floating networks of photosynthetic structures and trapping bladders. Interestingly, the generic GO term “lateral root development” was enriched among the *U. gibba*-expanded families (table 1 and [supplementary table S2, Supplementary Material](#) online). In total, three *U. gibba* genes belong to this functional category; two of them constitute ORTHOMCL18706 and another belongs to ORTHOMCL18829.

The *U. gibba* genes in ORTHOMCL18706 are likely homologous to the five highly conserved *Arabidopsis* genes *BREVIS RADIX (BRX)* and *BREVIS RADIX-like 1-4* (Mouchel et al. 2004; Briggs et al. 2006); *BRX* regulates root meristematic growth, possibly by controlling auxin response factor activity. ORTHOMCL18829 is composed of two *U. gibba* members encoding WRKY transcription factors, with only one of them annotated with the term lateral root development. Interestingly, when we used the corresponding sequences as queries to perform BLAST searches on the *Arabidopsis* genome, the best retrieved hit was *AtWRKY75 (AT5G13080)*, which plays a regulatory role in root hair patterning (Rishmawi et al. 2014). These two examples of enrichment for lateral root function may reflect gene functional specialization by co-option of orthologous genes during *U. gibba* evolution. The fact that *U. gibba* has no roots could suggest a heterotopic transfer of function of these genes to other organs, possibly ones bearing trichomes, which are developmentally related to root hairs. Indeed, bladder trichomes

are the sites of nutrient absorption in *U. gibba*, so it is tempting to speculate that some ancestral “root-specific” functions may have been transferred to traps. Indeed, both genes in ORTHOMCL18829 are preferentially expressed in bladders (Ibarra-Laclette et al. 2011). A potentially similar instance of evolutionary co-option may come from research on moss orthologs of the *Arabidopsis* bHLH transcription factors RHD6 and ATRSL1, reported to control root hair development. The corresponding orthologs in moss, PpRSL1 and PpRSL2, control the development of rhizoids, which are nonhomologous organs with a rooting function (Menand et al. 2007).

Furthermore, two genes annotated with the term “meristem maintenance,” likely homologous to *Arabidopsis* HOMEBOX GENE1 (*ATH1*, AT4G32980), were also found as enriched in expanded families. *ATH1* encodes a homeobox transcription factor involved in different photomorphogenic processes (Quaedvlieg et al. 1995), including modulation of growth at the interface between the stem, meristem, and organ primordia (Gómez-Mena and Sablowski 2008), and consequently, inflorescence architecture (Li et al. 2012).

#### The Putative Role of Gene Family Expansion in the Adaptation to Life in Aquatic Environments

*Utricularia gibba* lives in various habitats, including as a suspended aquatic or as a creeping plant in marshlands. It has been reported that the morphology of *Utricularia* species is plastic, and highly affected by environmental factors such as temperature, light, and water level (Taylor 1989; Reut and Fineran 2000). Accordingly, many GO term enrichments among *U. gibba*-expanded orthogroups are associated with response to environmental stimuli, including “response to abiotic stimulus,” several related to light (“response to light stimulus”—significant after correction, “response to blue light,” and “nonphotochemical quenching”), temperature (“response to temperature stimulus”), or salinity (“hypotonic salinity response”). The enrichments of these functional categories are consistent with key roles for various environmental stimuli on *U. gibba* development and physiology: *U. gibba* can live in very dynamic habitats that frequently transition among different environmental conditions. The rapid and efficient control of various metabolic and developmental responses to changing environmental constraints may therefore be critical to *U. gibba* survival.

For example, under typical environmental conditions, *U. gibba* produces normal, open bilateral flowers. However, under low light conditions, the lips are closed and the corollas do not open. Likewise, in deeper water or a submerged environment, the flowers lose corollas altogether and bear only closed calyces (Chao 2003). Both types of cleistogamous flowers are fertile and undergo self-fertilization. It is tempting to speculate that the GO term enrichment related to response to environmental stimuli may reflect the importance of fine-

tuned sensing of environmental factors participating in proper switching among flower-type developmental pathways.

In keeping with these unique flowering features, we discovered several enriched GO terms related to floral (meristem) development, such as “maintenance of floral meristem identity,” “meristem maintenance,” “floral whorl development,” “organ development,” “pollen tube development,” “adaxial/abaxial axis specification,” and “carpel development” (table 2 and [supplementary table S2, Supplementary Material](#) online). Among the genes annotated with the term maintenance of floral meristem identity, two clustered within orthogroup ORTHOMCL11149, an orthogroup comprising five *U. gibba* members likely homologous to *Arabidopsis* SHORT VEGETATIVE PHASE (*SVP*, AT2G22540). *SVP* encodes a MADS-box transcription factor that acts as a central regulator of flowering time (Hartmann et al. 2000). Although available expression data show only low to null expression in flowers (Ibarra-Laclette et al. 2011), the *U. gibba*-specific expansion of this family may be related to the critical function of *SVP*-like genes in the species’ flowering phenology.

Another highly significant enriched functional category (after correction) that may relate to the aquatic habit was “cuticle development.” Eleven of the 22 genes annotated with this term belong to the *U. gibba*-expanded orthogroup ORTHOMCL309, composed of 11 *U. gibba* members and 0–4 representatives from the other species. All of these genes are annotated as coding for 3-ketoacyl-CoA synthase enzymatic activity, involved in the biosynthesis of very long-chain fatty acids. The two *Arabidopsis* orthologs belonging to this orthogroup, AT1G25450 (*CER6*) and AT1G68530 (*CUT1*), catalyze the first step of the biosynthetic pathway for the layer of wax secreted onto the aerial surfaces of all land plants. This layer of wax protects plants from desiccation and other stresses, and allows survival in terrestrial environments (Millar et al. 1999; Fiebig et al. 2000). This hydrophobic cuticle layer evolved in land plants during the transition from an aquatic to a terrestrial environment (Go et al. 2014). The preferential retention and potential further functional diversification of duplicated genes from this family may reflect the metabolic flexibility required for waxy cuticle production in order to properly deal with the changing environments characteristic of *U. gibba*’s lifestyle. Accordingly, genes belonging to ORTHOMCL309 show broad and diverse expression patterns (Ibarra-Laclette et al. 2011).

#### Putative Roles of Adaptive Functional Specialization in *U. gibba* Phenotypic Diversification: A Case Study of Seven Candidate Genes

To further grasp the molecular evolutionary signatures underlying the evolution of *U. gibba*’s specialized body plan and unique lifestyle, we performed a preliminary genome-wide analysis of selection on the protein-coding regions of the 6,848 orthogroups containing homologs from all five species,



**Table 3**Summary of Genes Identified as under PS in *U. gibba*

Gene	<i>Arabidopsis</i> Ortholog	Function in <i>Arabidopsis</i>
<i>Scf00011.g1809.t1</i>	<i>AXR1</i>	A subunit of the RUB1-activating enzyme that regulates the protein degradation activity of Skp1-Cullin-Fbox complexes, which primarily, but not exclusively, affect auxin responses
<i>Scf00146.g10482.t1</i>	<i>UMAMIT41</i>	A nodulin <i>MtN21</i> -like transporter family protein that has a possible role in response to auxin in adventitious root formation
<i>Scf00260.g14017.t1</i>	<i>IGS</i>	Catalyzes the fourth step of the tryptophan biosynthesis pathway, and is as such responsible for the synthesis of indole-3-glycerolphosphate, the intermediate serving as a branchpoint to the Trp-independent pathway for auxin biosynthesis
<i>Scf00083.g7570.t1</i>	<i>TAR2</i>	Involved in the indole-3-pyruvic acid (IPyA) pathway, one of the Trp-dependent pathways for auxin biosynthesis. Together with <i>TAA1</i> , affects root meristem maintenance and differential growth in apical hooks
<i>Scf00505.g18737.t1</i>	<i>SOL1</i>	A suppressor of root-specific overexpression of <i>CLE19</i> . The <i>sol1</i> mutant partially suppresses the loss of root meristem maintenance and short root phenotype caused by <i>CLE19</i> overexpression
<i>Scf00003.g611.t1</i>	<i>DEG9</i>	The only Deg protease member located in the nucleus. Hypothesized to have functions involving ribosome-related transcription and modification.
<i>Scf00324.g15492.t1</i>	<i>DEG10</i>	Highly induced after treatment with inhibitors of the mitochondrial electron transport chain, which suggests a role of DEG10 in protein quality control via degradation of damaged mitochondrial proteins

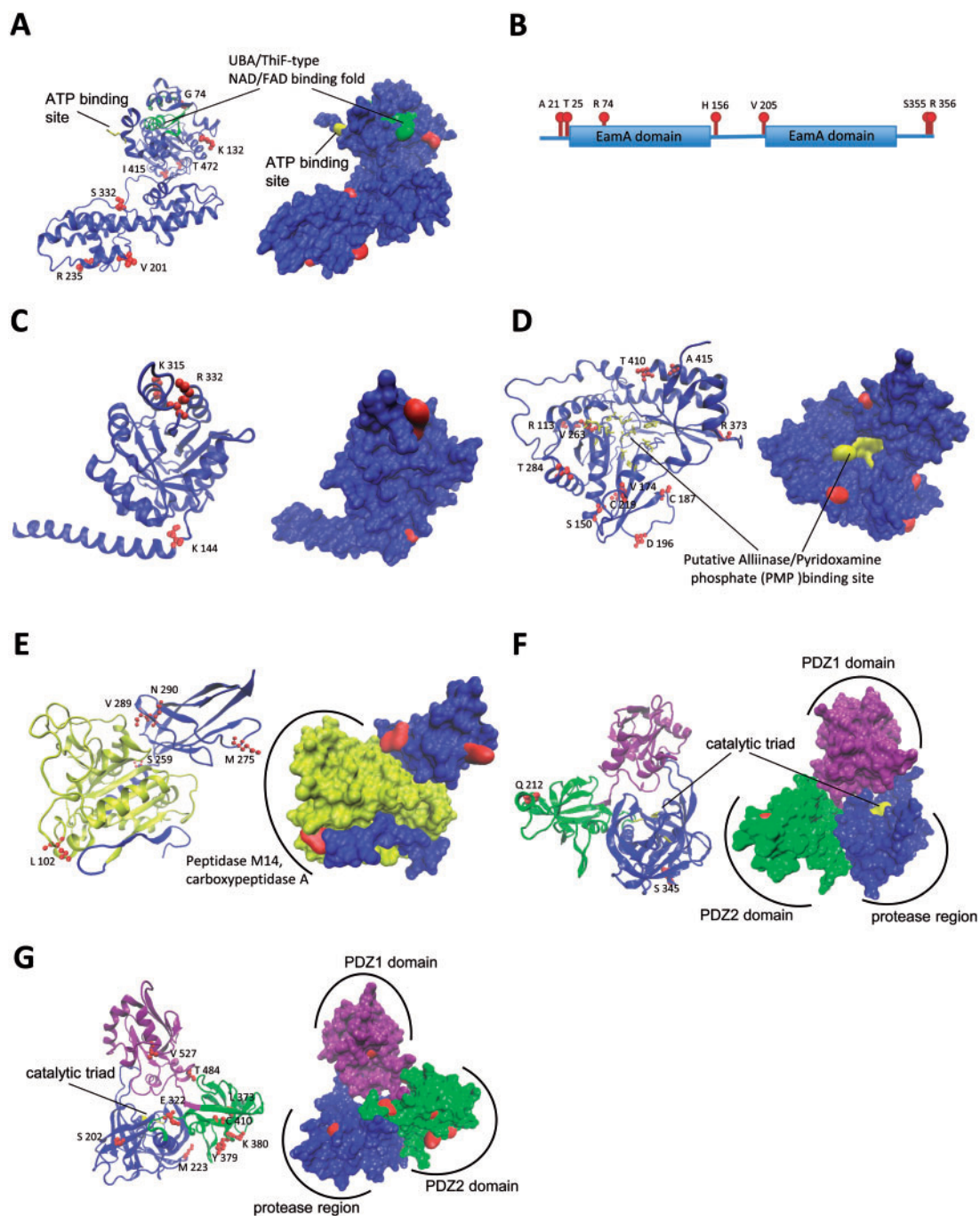
and that exceeded the quality criteria filters (see Materials and Methods; [supplementary table S6, Supplementary Material](#) online). Different codon-substitution evolutionary models assessing for asymmetric evolution, divergent selective pressures or PS specifically occurring in the *U. gibba* ortholog were implemented in PAML (Yang 1997, 2007). From these preliminary results, we selected seven candidate genes that showed the molecular hallmarks of PS in the *U. gibba* lineage. These seven cases were subjected to manual curation and further examination for PS using a greater sampling of orthologous sequences (table 3).

First, we focused on four genes involved in several critical steps of auxin metabolism, signaling of primary root abortion, and vegetative meristem growth, as these may have played a major role in the evolution of *U. gibba*'s specialized body plan. For example, *Scf00011.g1809.t1* is orthologous to *Arabidopsis AXR1* (At1G05180; [supplementary fig. S1, Supplementary Material](#) online), a gene regulating the protein degradation activity of Skp1-Cullin-Fbox complexes, which primarily affect auxin responses (Leyser et al. 1993; del Pozo et al. 2002). Similarly, *Scf00146.g10482.t1* is orthologous to *Arabidopsis UMAMIT41* (At3G28050; [supplementary fig. S2, Supplementary Material](#) online), which is a nodulin *MtN21*-like transporter family protein that may respond to IAA in adventitious root formation (Busov et al. 2004). Beside these two auxin/IAA signaling related genes, we also selected *Scf00260.g14017.t1*, an ortholog of *indole-3-glycerol phosphate synthase (IGS)* (AT2G04400; [supplementary fig. S3, Supplementary Material](#) online), partly responsible for the synthesis of indole-3-glycerolphosphate, which is the intermediate serving as a branch point to the Trp-independent

pathway for auxin synthesis (Tzin and Galili 2010). Furthermore, another gene identified as subjected to PS in *U. gibba* was *Scf00083.g7570.t1*, an ortholog of *tryptophan aminotransferase related 2 (TAR2)*; [supplementary fig. S4, Supplementary Material](#) online), involved in the indole-3-pyruvic acid (IPyA) pathway, one of the Trp-dependent pathways for IAA biosynthesis. By generating auxin gradients among different parts of plant tissue, *TAR2* and *TAA1* affect root meristem maintenance and differential growth in apical hooks (Stepanova et al. 2008). Therefore, our analysis suggests an evolutionary fine-tuning of genes involved in inhibiting primary root elongation, which can stimulate the initiation of lateral and adventitious roots (Vanneste and Friml 2009).

Another gene found to be evolving under PS in *U. gibba* was *Scf00505.g18737.t1*, an ortholog of *Arabidopsis SOL1* (AT1G71696; [supplementary fig. S5, Supplementary Material](#) online). *SOL1* was isolated as a suppressor of root-specific overexpression of *CLE19*, a *CLAVATA3*-like gene (Casamitjana-Martinez et al. 2003). Because the *sol1* mutant partially suppresses the loss of root meristem maintenance and short root phenotype caused by *CLE19* overexpression, the signature of PS found in the *SOL1* ortholog might reflect a similar function related to arrested root differentiation.

Finally, two additional *U. gibba* genes examined were *Scf00003.g611.t1* and *Scf00324.g15492.t1*, orthologs of *Arabidopsis DEG9* (AT5G40200) and *DEG10* (AT5G36950), respectively ([supplementary figs. S6 and S7, Supplementary Material](#) online). According to the MEROPS database classification (<http://merops.sanger.ac.uk/>), *DEG9* and *DEG10* belong to the peptidase family S1 (chymotrypsin family),



**FIG. 2.**—Protein 3D-structure models and architecture of functional domains of seven *Utricularia gibba* genes. (A and C–G) Cartoon backbones highlighting secondary structures (left side of the panel) and molecular surface representations (right side of the panel) for *U. gibba* AXR1, IGS, TAR2, SOL1, DEG9, and DEG10, respectively. The 3D-structural models were obtained using SWISS MODEL. (B) Architecture of protein functional domains of *U. gibba* UMAMIT41 as retrieved from INTERPROSCAN. Putative PS residues are shown in red. Functionally relevant residues or protein domains are shown in different colors.

which is widely distributed throughout all kingdoms of life. In animals, these peptidases play many different roles, including intestinal digestion and IgA-mediated immune responses (Lister et al. 2004). The functions of these peptidases in plants are not well known. *Arabidopsis* DEG9 was found as

the only Deg protease member to be located in the nucleus, and was hypothesized to have functions involving ribosome-related transcription and modification. *DEG10* was reported to be highly induced after treatment of *Arabidopsis* plants with inhibitors of the mitochondrial electron transport chain,

which suggested a role of DEG10 in protein quality control via degradation of damaged mitochondrial proteins (Lister et al. 2004). Apart from the tempting hypothesis that they may also play a role in prey digestion, these Deg proteases may provide essential protection against damage to cellular proteins, for example, as incurred by ROS.

Many of the PS amino acid sites found in *U. gibba* orthologous proteins occurred in positions well conserved in the remaining sequences included in alignments, and some involved radical changes in the physicochemical properties of the amino acids (fig. 2 and supplementary figs. S1–S7, Supplementary Material online). To gain further insights into the putative functional roles of the amino acid changes fixed by PS, we mapped them onto 3D-structural models when available, or the protein functional domains as predicted by INTERPROSCAN (Zdobnov and Apweiler 2001). In each case, functionally significant residues were identified from the literature or predicted by Evolutionary Trace (Lichtarge et al. 1996; fig. 2). Interestingly, some of the PS amino acid site changes were located within functional domains or residues, suggestive of their possible biological significance. For example, two of the residues (174 V and 187 C) inferred to be subjected to PS in the *U. gibba* TAR2 ortholog were within the protein's surface patch (Tao et al. 2008; fig. 2D). In addition, in the *U. gibba* DEG9 ortholog, one PS residue (212 Q) was located in the PDZ2 domain, whereas another (345 S) occurred in the protease domain (fig. 2F). In *U. gibba* DEG10 orthologs, five PS residues (322 E, 373 L, 379 Y, 380 K, and 410 C) were in the PDZ2 domain, whereas two additional PS residues (202 S and 223 M) lay in the protease region (fig. 2G). Since the PDZ2 domain mediates hexamer formation and locks the protease into the resting state (Sun et al. 2012), its PS amino acid changes may modify their possible coordination in specific protein functions.

## Conclusions

We found examples of both lineage-specific gene family expansions and putative adaptive functional specialization of *U. gibba* orthologs that could have contributed to *U. gibba* evolution. Our genome-wide analyses of gene family expansions and protein-coding gene evolution provide useful information to explore possible mechanisms underlying both morphological and physiological specializations of *U. gibba*. Our results further suggest interesting target genes and gene families for future experimental confirmation to validate their possible adaptive functional roles in *U. gibba*'s highly specialized body plan and unique carnivorous lifestyle.

## Supplementary Material

Supplementary tables S1–S7, figures S1–S7, and files S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>)

## Acknowledgments

This study was supported in part by the United States National Science Foundation, award 0922742 to V.A.A.

## Literature Cited

- Adamec L. 2006. Respiration and photosynthesis of bladders and leaves of aquatic *Utricularia* species. *Plant Biol.* 8:765–769.
- Adlassnig W, Peroutka M, Lambers H, Lichtscheidl IK. 2005. The roots of carnivorous plants. *Plant Soil* 274:127–140.
- Albert VA, Jobson RW, Michael TP, Taylor DJ. 2010. The carnivorous bladderwort (*Utricularia*, Lentibulariaceae): a system inflates. *J Exp Bot.* 61:5–9.
- Aloy P, et al. 2001. The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *J Biol Chem.* 276:16177–16184.
- Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342:1241089.
- Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 57:289–300.
- Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol.* 59:121–132.
- Birney E, Durbin R. 2000. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* 10:547–548.
- Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678.
- Bolwell GP. 1999. Role of active oxygen species and NO in plant defence responses. *Curr Opin Plant Biol.* 2:287–294.
- Briggs GC, Mouchel CF, Hardtke CS. 2006. Characterization of the plant-specific BREVIS RADIX gene family reveals limited genetic redundancy despite high sequence conservation. *Plant Physiol.* 140:1306–1316.
- Busov VB, et al. 2004. An auxin-inducible gene from loblolly pine (*Pinus taeda* L.) is differentially expressed in mature and juvenile-phase shoots and encodes a putative transmembrane protein. *Planta* 218:916–927.
- Carretero-Paulet L, Fares MA. 2012. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Mol Biol Evol.* 29:3541–3551.
- Casamitjana-Martinez E, et al. 2003. Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of *Arabidopsis* root meristem maintenance. *Curr Biol.* 13:1435–1441.
- Chao Y-S. 2003. A Taxonomic study on Lentibulariaceae of Taiwan [PhD Thesis]. Department of Biological Sciences, National Sun Yat-sen University, Taiwan.
- Cho SK, Ryu MY, Song C, Kwak JM, Kim WT. 2008. *Arabidopsis* PUB22 and PUB23 are homologous U-box E3 ubiquitin ligases that play combinatory roles in response to drought stress. *Plant Cell* 20:1899–1914.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Cui L, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16:738–749.
- Darwin C. 1875. *Insectivorous plants*. New York: D. Appleton and Co.
- De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol Evol.* 20:591–597.

- del Pozo JC, et al. 2002. AXR1-ECR1-dependent conjugation of RUB1 to the *Arabidopsis* cullin AtCUL1 is required for auxin response. *Plant Cell* 14:421–433.
- Denoeud F, et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345: 1181–1184.
- Ellison AM, Gotelli NJ. 2001. Evolutionary ecology of carnivorous plants. *Trends Ecol Evol.* 16:623–629.
- Ellison AM, Gotelli NJ. 2009. Energetics and the evolution of carnivorous plants—Darwin's 'most wonderful plants in the world'. *J Exp Bot.* 60: 19–42.
- Fei Z, Tang X, Alba R, Giovannoni J. 2006. Tomato expression database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Res.* 34:D766–D770.
- Fiebig A, et al. 2000. Alterations in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* 12:2001–2008.
- Givnish TJ. 1989. Ecology and evolution of carnivorous plants. *Plant-Animal Interact.* 243:479–497.
- Go YS, Kim H, Kim HJ, Suh MC. 2014. *Arabidopsis* cuticular wax biosynthesis is negatively regulated by the DEWAX gene encoding an AP2/ERF-type transcription factor. *Plant Cell* 26:1666–1680.
- Gómez-Mena C, Sablowski R. 2008. ARABIDOPSIS THALIANA HOMEBOX GENE1 establishes the basal boundaries of shoot organs and controls stem growth. *Plant Cell* 20:2059–2072.
- Gu Z, et al. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421:63–66.
- Hartmann U, et al. 2000. Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *Plant J.* 21:351–360.
- Hatano N, Hamada T. 2008. Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *J Proteomics Res.* 7:809–816.
- Hedges S. B., Dudley J., Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23), 2971–2972.
- Huang DT, et al. 2007. Basis for a ubiquitin-like protein thioester switch toggling E1–E2 affinity. *Nature* 445:394–398.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph.* 14:33–38, 27–38.
- Ibarra-Laclette E, et al. 2011. Transcriptomics and molecular evolutionary rate analysis of the bladderwort (*Utricularia*), a carnivorous plant with a minimal genome. *BMC Plant Biol.* 11:101.
- Ibarra-Laclette E, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498:94–98.
- Jiao Y, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100.
- Juniper B, Robins R, Joel D. 1989. The carnivorous plants. London: Academic Press.
- Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 110:462–467.
- Kaku H, et al. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc Natl Acad Sci U S A.* 103:11086–11091.
- Laakkonen L, Jobson R, Albert V. 2006. A new model for the evolution of carnivory in the bladderwort plant (*Utricularia*): adaptive changes in cytochrome c oxidase (COX) provide respiratory power. *Plant Biol.* 8: 758–764.
- Leyser HO, et al. 1993. *Arabidopsis* auxin-resistance gene AXR1 encodes a protein related to ubiquitin-activating enzyme E1. *Nature* 364: 161–164.
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of orthology groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Li WH, Gojobori T, Nei M. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* 292:237–239.
- Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. *Nature* 409:847–849.
- Li Y, Pi L, Huang H, Xu L. 2012. ATH1 and KNAT2 proteins act together in regulation of plant inflorescence architecture. *J Exp Bot.* 63: 1423–1433.
- Librado P, Vieira FG, Rozas J. 2012. BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28:279–281.
- Lichtarge O, Bourne HR, Cohen FE. 1996. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol.* 257: 342–358.
- Lister R, et al. 2004. A transcriptomic and proteomic characterization of the *Arabidopsis* mitochondrial protein import apparatus and its response to mitochondrial dysfunction. *Plant Physiol.* 134:777–789.
- Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.* 102:10557–10562.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Menand B, et al. 2007. An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316:1477–1480.
- Millar AA, et al. 1999. CUT1, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* 11:825–838.
- Mithofer A. 2011. Carnivorous pitcher plants: insights in an old topic. *Phytochemistry* 72:1678–1682.
- Moore RC, Purugganan MD. 2003. The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A.* 100:15682–15687.
- Mouchel CF, Briggs GC, Hardtke CS. 2004. Natural genetic variation in *Arabidopsis* identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev.* 18:700–714.
- Muller K, et al. 2004. Evolution of carnivory in Lentibulariaceae and the Lamiales. *Plant Biol.* 6:477–490.
- Muller KF, Borsch T, Legendre L, Porembski S, Barthlott W. 2006. Recent progress in understanding the evolution of carnivorous lentibulariaceae (lamiales). *Plant Biol.* 8:748–757.
- Myhre S, Tveit H, Mollestad T, Laegreid A. 2006. Additional gene ontology structure for improved biological reasoning. *Bioinformatics* 22: 2020–2027.
- Nei M. 2005. Selectionism and neutralism in molecular evolution. *Mol Biol Evol.* 22:2318–2342.
- Ogata H, et al. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27:29–34.
- Ohno S. 1970. Evolution by gene duplication. London: George Allen & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.
- Penn O, et al. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic acids research* 38:W23–W28.
- Plachno BJ, et al. 2006. Fluorescence labelling of phosphatase activity in digestive glands of carnivorous plants. *Plant Biol.* 8:813–820.
- Proost S, Pattyn P, Gerats T, Van de Peer Y. 2011. Journey through the past: 150 million years of plant genome evolution. *Plant J.* 66: 58–65.
- Quaedvlieg N, Dockx J, Rook F, Weisbeek P, Smeekens S. 1995. The homeobox gene ATH1 of *Arabidopsis* is derepressed in the photomorphogenic mutants cop1 and det1. *Plant Cell* 7:117–129.
- Renner T, Specht CD. 2012. Molecular and functional evolution of class I chitinases for plant carnivory in the Caryophyllales. *Mol Biol Evol.* 29: 2971–2985.
- Renner T, Specht CD. 2013. Inside the trap: gland morphologies, digestive enzymes, and the evolution of plant carnivory in the Caryophyllales. *Curr Opin Plant Biol.* 16:436–442.
- Reut MS, Fineran BA. 2000. Ecology and vegetative morphology of the carnivorous plant *Utricularia dichotoma* (Lentibulariaceae) in New Zealand. *N Z J Bot.* 38:433–450.

- Rishmawi L, et al. 2014. Non-cell-autonomous regulation of root hair patterning genes by WRKY75 in *Arabidopsis*. *Plant Physiol.* 165: 186–195.
- Rottloff S, et al. 2011. Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. *J Exp Bot.* 62:4639–4647.
- Salzer P, et al. 2000. Differential expression of eight chitinase genes in *Medicago truncatula* roots during mycorrhiza formation, nodulation, and pathogen infection. *Mol Plant Microbe Interact.* 13: 763–777.
- Shibuya N, Minami E. 2001. Oligosaccharide signalling for defence responses in plant. *Physiol Mol Plant Pathol.* 59:223–233.
- Sirová D, Adamec L, Vrba J. 2003. Enzymatic activities in traps of four aquatic species of the carnivorous genus *Utricularia*. *New Phytol.* 159: 669–675.
- Soltis DE, Burleigh JG. 2009. Surviving the K-T mass extinction: new perspectives of polyploidization in angiosperms. *Proc Natl Acad Sci U S A.* 106:5455–5456.
- Stepanova AN, et al. 2008. TAA1-Mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* 133:177–191.
- Sun R, et al. 2012. Crystal structure of *Arabidopsis* Deg2 protein reveals an internal PDZ ligand locking the hexameric resting state. *J Biol Chem.* 287:37564–37569.
- Tao Y, et al. 2008. Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. *Cell* 133: 164–176.
- Taylor P. 1989. The genus *Utricularia*: a taxonomic monograph. *Kew Bull. Add. Ser.* 14.
- Trujillo M, Ichimura K, Casais C, Shirasu K. 2008. Negative regulation of PAMP-triggered immunity by an E3 ubiquitin ligase triplet in *Arabidopsis*. *Curr Biol.* 18:1396–1401.
- Turrens JF. 2003. Mitochondrial formation of reactive oxygen species. *J Physiol.* 552:335–344.
- Tzin V, Galili G. 2010. The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *The Arabidopsis book/American Society of Plant Biologists* 8.
- Vanneste S, Friml J. 2009. Auxin: a trigger for change in plant development. *Cell* 136:1005–1016.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15:568–573.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19:908–917.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22:1107–1118.
- Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.

Associate editor: Laura Landweber