

# Combination of transcriptomic and proteomic approaches helps to unravel the protein composition of *Chelidonium majus* L. milky sap

Robert Nawrot<sup>1</sup> · Jakub Barylski<sup>1</sup> · Rico Lippmann<sup>2,3</sup> · Lothar Altschmied<sup>2</sup> · Hans-Peter Mock<sup>2</sup>

Received: 30 January 2016 / Accepted: 1 July 2016 / Published online: 11 July 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

## Abstract

**Main conclusion** A novel annotated *Chelidonium majus* L. transcriptome database composed of 23,004 unique coding sequences allowed to significantly improve the sensitivity of proteomic *C. majus* assessments, which showed novel defense-related proteins characteristic to its latex.

To date, the composition of *Chelidonium majus* L. milky sap and biosynthesis of its components are poorly characterized. We, therefore, performed *de novo* sequencing and assembly of *C. majus* transcriptome using Illumina technology. Approximately, 119 Mb of raw sequence data was obtained. Assembly resulted in 107,088 contigs, with N50 of 1913 bp and N90 of 450 bp. Among 34,965 unique coding sequences (CDS), 23,004 obtained CDS database served as a basis for further proteomic analyses. The database was then used for the identification of proteins from *C. majus* milky sap, and whole plant extracts analyzed using liquid chromatography–electrospray ionization–tandem mass spectrometry (LC–ESI–MS/MS) approach. Of about 334 different putative proteins were identified in *C. majus*

milky sap and 1155 in *C. majus* whole plant extract. The quantitative comparative analysis confirmed that *C. majus* latex contains proteins connected with response to stress conditions and generation of precursor metabolites and energy. Notable proteins characteristic to latex include major latex protein (MLP, presumably belonging to Bet v1-like superfamily), polyphenol oxidase (PPO, which could be responsible for browning of the sap after exposure to air), and enzymes responsible for anthocyanidin, phenylpropanoid, and alkaloid biosynthesis.

**Keywords** Greater celandine · Latex · Laticifers · Major latex protein · Papaveraceae · Polyphenol oxidase · Sequencing

## Abbreviations

CDS Coding sequences  
GO Gene ontology

## Introduction

Greater celandine (*Chelidonium majus* L.) (Fig. 1), an herbaceous perennial plant, belongs to Papaveraceae family, which is an important source of biologically active substances. Milky sap and extracts of greater celandine are used in traditional medicine to treat papillae, warts, and condylomas, which are symptoms of human papilloma virus (HPV) infections. *C. majus* extracts are also used to treat liver disorders and fight fever (Hiller et al. 1998). The medicinal and pharmaceutical interest in this plant is based on its ability to synthesize alkaloids, flavonoids, or phenolic acids (Colombo and Bosisio 1996). It is often suspected that many of these substances may be either synthesized or stored in laticifers of the plant. These

**Electronic supplementary material** The online version of this article (doi:10.1007/s00425-016-2566-7) contains supplementary material, which is available to authorized users.

✉ Robert Nawrot  
rnawrot@amu.edu.pl

<sup>1</sup> Department of Molecular Virology, Institute of Experimental Biology, Faculty of Biology, Adam Mickiewicz University in Poznań, Umultowska 89, 61-614 Poznań, Poland

<sup>2</sup> Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany

<sup>3</sup> Present Address: Sandoz GmbH, Biochemiestraße 10, 6250 Kundl, Austria



**Fig. 1** *Chelidonium majus* L. plant. **a** Adult plant grown in natural habitat during flowering. **b** Yellow flowers. **c** 6-week-old stem of the plant with exuding yellow milky sap, which was cut and used for RNA isolation

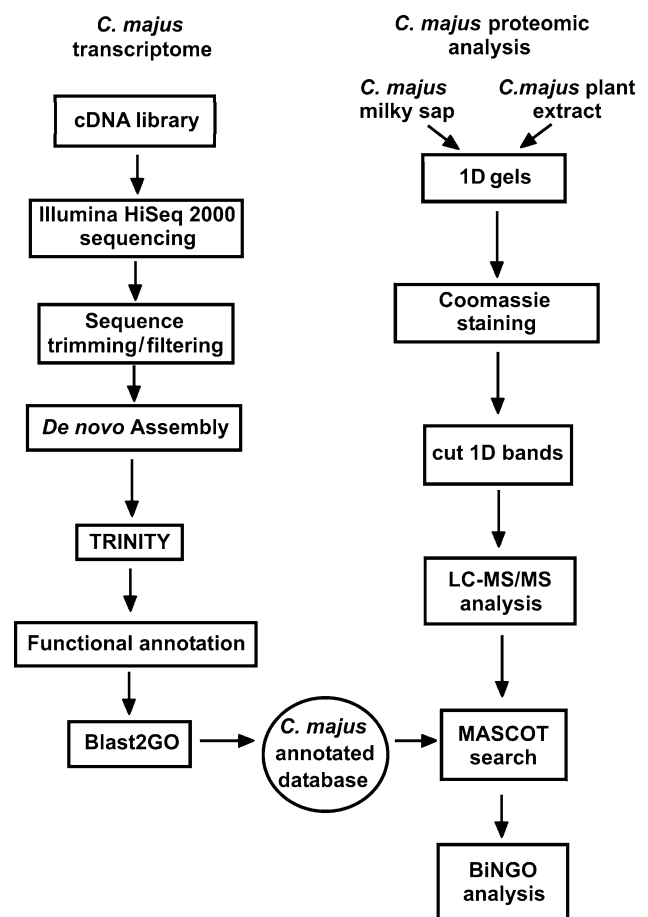
specialized cells produce milky sap (latex) that exudes when the plant is injured (Hagel et al. 2008). They are located in phloem areas forming an internal, articulated, and non-anastomosing system throughout the whole plant (Hagel et al. 2008). Therefore, sequestration of bioactive compounds into laticifers might protect the plant from the effects of its own toxins and provide a defense against herbivores (Hagel et al. 2008; Konno 2011; Escalante-Perez et al. 2012).

Previous proteomic studies regarding *C. majus* milky sap using two-dimensional electrophoresis (2-DE) and liquid chromatography–electrospray ionization–tandem mass spectrometry (LC–ESI–MS/MS) revealed the presence of about 21 proteins, mostly of defense- and pathogenesis-related properties (Nawrot et al. 2007a). However, sensitivity of used methods suffered from lack of relevant database of reference sequences.

*Chelidonium majus* (Fig. 1) belongs to non-model plants, and to date, its genome is unsequenced. Since assembly of such genome could be very laborious and correct gene annotation would require either homology-based or transcriptomic information, we decided to use RNA sequencing to create such database. Recent advances in next-generation sequencing (NGS) with the possibility of quantification of transcript abundances have greatly lowered its cost and provided insight into diverse plant species, including those with medicinal properties (Hao et al. 2011; Hamilton and Buell 2012). The *de novo* transcriptome sequencing and characterization has been performed successfully for *Taxus wallichiana* var. *mairei* (Lemée & H.Lév.) L.K.Fu & Nan Li (Hao et al. 2011), *Phalaenopsis* spp. (orchid) (Fu et al. 2011), *Macleaya* spp.

(plume poppies) (Zeng et al. 2013), *Beta vulgaris* L. (sugarbeet) (Fugate et al. 2014), *Solanum tuberosum* L. (potato) (Gong et al. 2015), *Panax ginseng* C.A.Mey. (Jayakodi et al. 2015), *Salvia miltiorrhiza* Bunge (Xu et al. 2015), and many others.

The aim of the present study was to sequence, assemble, and annotate the *C. majus* transcriptome. We created a database for protein identification for mass spectrometry data. The database allowed us to significantly improve the sensitivity of previous assessments concerning composition of *C. majus* milky sap, which relied only on NCBI nr (National Center for Biotechnology Information non-redundant) database searches with *Viridiplantae* filter (Nawrot et al. 2007a, b, 2014). A schematic representation of the overall sequencing, annotation, and protein identification workflow is presented in Fig. 2. The integration of transcriptomic and proteomic data in this study shows that it is possible to answer the emerging biological issues on non-model plant species without genome sequence available, such as novel insights into plant latex protein composition.



**Fig. 2** Schematic representation of the overall sequencing and annotation workflow for *C. majus* transcriptome together with proteomic analysis

## Materials and methods

### Plant material

*Chelidonium majus* L. plants were planted in IPK Gatersleben green house from *C. majus* seeds collected in Poznań, Poland, on 12 June 2012. The voucher specimen of seeds is deposited in the Department of Molecular Virology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poland. They were cultivated in small cuvette with pH 5.5 ground, and after 4 weeks, they were transferred to bigger pots. Supplemented with mineral substrates (N 6 g/m<sup>2</sup>, P<sub>2</sub>O<sub>5</sub> 7 g/m<sup>2</sup>, K<sub>2</sub>O 9 g/m<sup>2</sup>), they were cultivated until reaching the height of ca. 25–30 cm for 6 weeks.

### Illumina HiSeq 2000 sequencing

*Chelidonium majus* RNA samples were isolated from stem of the plant with exuding milky sap (Fig. 1) using RNeasy Plant Mini Kit (Qiagen). cDNA synthesis was carried out and cDNA library was prepared using TruSeq RNA Library Prep Kit v2 (Illumina Inc., San Diego, CA, USA). After RNA quality validation on Agilent Technologies 2100 Bioanalyzer, the library was sequenced using Illumina HiSeqTM 2000 (Illumina Inc.).

### De novo transcriptome assembly

Paired RNA-seq reads were pre-trimmed at the first undetermined base or at the first base having Phred quality below 20. The pairs with one (or both) reads shorter than 31 bases after trimming were excluded from the assembly process. In total, 112,544,804 (84 %) out of 133,550,365 read pairs passed filtering. Subsequently, the transcripts were assembled *de novo* using Trinity r2013-02-25 (specifically designed to deal with *de novo* transcriptome assembly, with default settings) (Grabherr et al. 2011). Protein coding sequences (CDS) were predicted using TransDecoder (<http://transdecoder.sourceforge.net/>) from Trinity package. Non-redundant CDS were selected for annotation and further analyses. *De novo* assembly was performed by VitaInSilica (<http://www.vitainsilica.pl/>).

### Gene annotation and analysis

To obtain comprehensive information on the function of predicted CDS, we used Blast2GO that annotates sequences based on ontological data retrieved for top BLAST (Basic Local Alignment Search Tool) hits of each sequence. To generate the comparison, local BLASTx search with the e value cutoff of 1e-10 against complete RefSeq (NCBI Reference Sequence) protein database

(release 64, available at NCBI FTP server from 14 March 2014) was performed. The resulting XML document was imported to Blast2GO software and refined by imposing more stringent cutoff (1e-25). After InterProScan analysis, the program assigned provisional names and gene ontology (GO) terms to annotated CDS. The additional BLAST analysis performed during annotation step allowed us to scrutinize sequences for remaining contaminants of non-plant origin. Additionally, KAAS [Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server] was used to functionally annotate genes in a genome by amino acid sequence comparisons against a manually curated set of ortholog groups from dicot plants in KEGG GENES (<http://www.genome.jp/kegg/genes.html>).

### One-dimensional gel electrophoresis (1-D, SDS-PAGE)

To verify the protein composition of protein samples, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in a slab mini-gel apparatus according to Laemmli (1970), using 10 % polyacrylamide as the separating gel and 5 % polyacrylamide as the stacking gel. The proteins were reduced by heating them to 100 °C in the presence of 2-mercaptoethanol for 5 min. About 50 µg of each sample was added to the gel, including two technical replicates. After SDS-PAGE, the gels were fixed and stained using sensitive Coomassie blue staining (Neuhoff et al. 1988).

### Liquid chromatography and tandem mass spectrometry analysis (LC–MS/MS)

Gel bands were digested with trypsin and, finally, all were subjected to LC–ESI-MS/MS analysis in Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland. Tryptic peptide mixtures were analyzed by LC–MS/MS using nanoflow HPLC and a Linear Trap Quadrupole (LTQ) Orbitrap XL mass spectrometer (Thermo Fisher Scientific) as the mass analyzer. Peptides were eluted from a 75-µm analytical column on a linear gradient from 10 to 30 % acetonitrile over 50 min and sprayed directly into the LTQ-Orbitrap mass spectrometer.

### Tandem mass spectrometry (MS/MS) data analysis

Proteins were identified by MS/MS via information-dependent acquisition of fragmentation spectra (Nawrot et al. 2007a, 2014) using MASCOT 2.4.1 search (Perkins et al. 1999; Matrix Science, London, UK; [www.matrixscience.com](http://www.matrixscience.com)) against *C. majus* CDS database.

## Quantitative comparative analysis of *C. majus* milky sap proteins

For quantitative, comparative analysis of *C. majus* milky sap with *C. majus* whole plant samples based on emPAI (exponentially modified protein abundance index) values of individual proteins, we analyzed data from three independent specimens of latex and three corresponding whole plant extracts. emPAI value is proportional to protein abundance in a protein mixture (Ishihama et al. 2005). Approximate relative abundance of each protein (calculated from emPAI according to formula: (emPAI value  $\times$  100)/sum of all emPAI values in the category) was averaged among three latex replicates and compared with the corresponding value calculated for whole plant extract. The results with a sap/extract ratio  $>2$  and *t* test significance level  $<0.1$  were considered as significantly overrepresented in the sap.

## Biological networks gene ontology tool (BiNGO) analysis

The overrepresentation of ontological terms in tested set (milky sap) compared to whole plant extract proteome and transcriptome was assessed using BiNGO tool. BiNGO (Maere et al. 2005) is a Cytoscape (Saito et al. 2012) plugin determining which GO categories are statistically over- or underrepresented in a set of genes. It maps the predominant functional themes of a given gene set on the GO hierarchy and outputs this mapping as a Cytoscape graph.

## Results

### Illumina sequencing and *de novo* *C. majus* transcriptome assembly

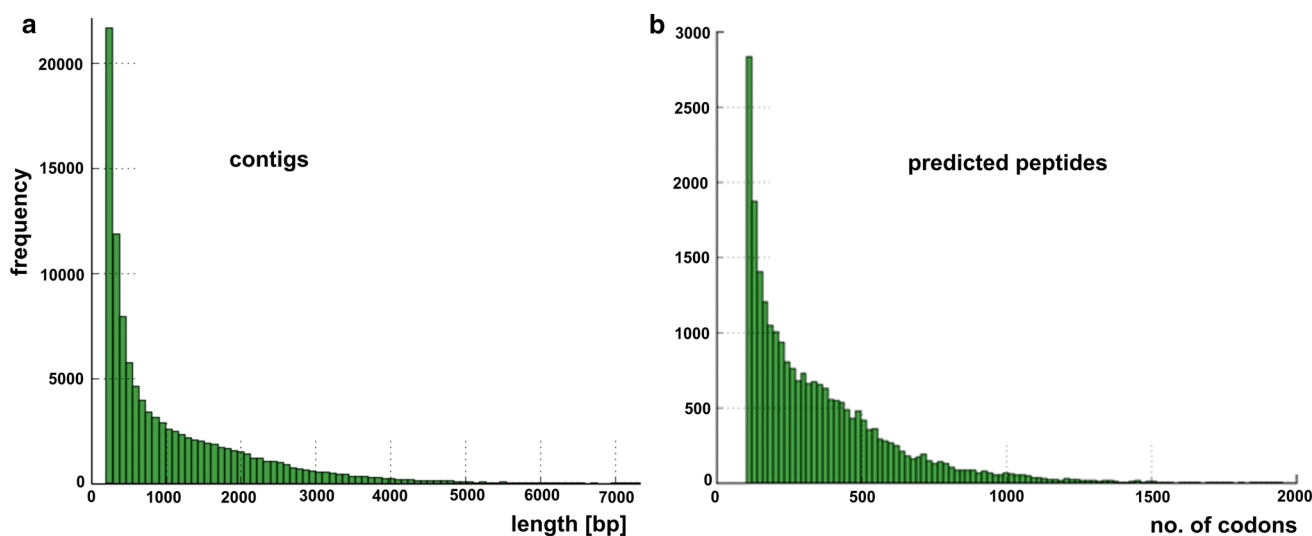
With the purpose of understanding *C. majus* transcriptome, RNA was extracted from the young stem of 6-week-old *C. majus* plant with exuding milky sap (Fig. 1c) and sequenced with Illumina paired-end sequencing technology generating approximately 119 Mb of raw sequence data from 133,550,365 reads with an average length of almost 100 bp (Fig. 3; Table 1).

After the quality and adaptor trimming process, *de novo* transcriptome assembly was performed using Trinity suit (specifically designed to deal with *de novo* transcriptome assembly) (Grabherr et al. 2011). This resulted in 107,088 contigs, with N50 of 1913 bp and N90 of 450 bp (Table 1).

### Annotation of *C. majus* transcriptome

TransDecoder detected 52,896 CDS including 35,271 (66.68 %) complete ones (Table 1). In addition, 9137 RNAs with length over 1 kb with no coding potential (lncRNAs) were assembled suggesting significant presence.

Detailed analysis of BLAST results revealed that 33 of coding sequences (0.09 %) may be contaminants of either fungal or bacterial origin (likely transcripts of plant commensal flora). These sequences were removed from the



**Fig. 3** *Chelidonium majus* transcriptome assembly overview. **a** Length distributions of assembled contigs. **b** Length distributions of predicted peptides (peptide lengths for raw transcripts): mean

356.44, standard deviation 267.36. Data obtained after Illumina paired-end sequencing of the normalized 6-week-old stem cDNA library

**Table 1** Summary of *C. majus* transcriptome assembly and annotation

Assembly		Annotation	
Contigs	107,088	CDS	52,896
Cumulative assembly size	119,198,595	Complete	35,271
GC (%)	39.27	3'-partial	4182
Contigs >1 kb	41,454	5'-partial	9914
Bases in contigs >1 kb	90,034,432	Internal	3529
N50	1913	Unique	34,965
N90	450	Contaminant	33
Number of gaps	0	Remaining	34,932
Longest	15,519	Annotated	23,004

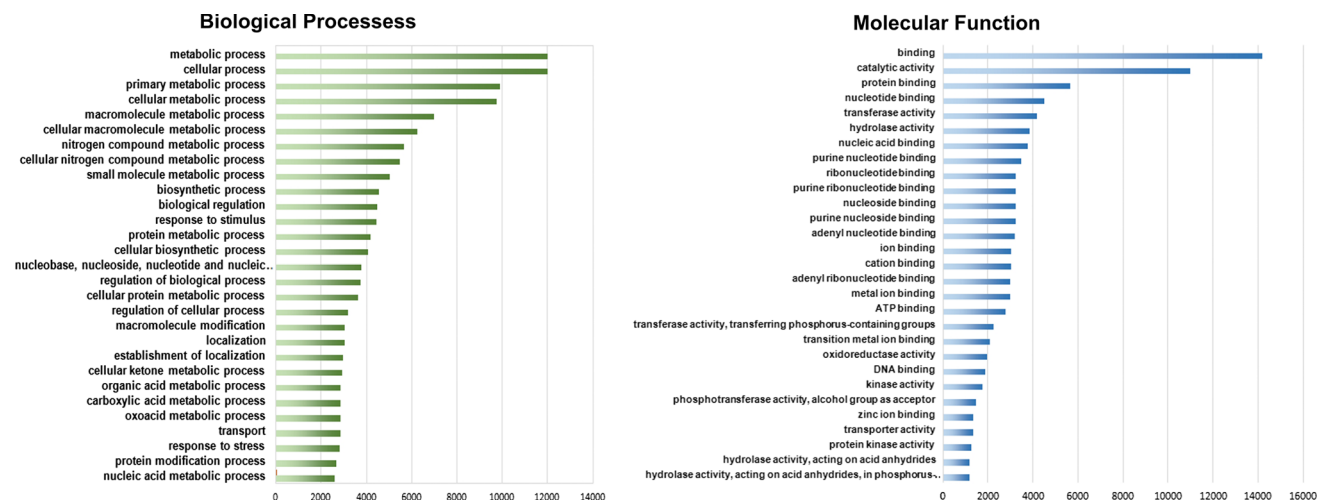
N50 and N90 define the length of contig for which the set of contigs with greater or equal length corresponds to 50 and 90 % of the assembly, respectively. CDS were classified as complete (both start and stop codon present), 3'-partial (only stop codon), 5'-partial (only start codon), and internal (neither start nor stop codon present) CDS. Unique CDS are all non-redundant CDS (regardless of completeness) detected during the study

analyzed set. About 89.07 % of the remaining unique CDS had at least one BLAST hit against RefSeq database if a value cutoff equal to 1e-10 was applied. Proportion dropped to 82.09 % when more stringent cutoff (1e-25) was imposed. InterProScan utility found protein domains in 89.32 % CDS. About 35.98 % (12,692) sequences were annotated using KAAS with KO (KEGG Orthology) assignments and automatically generated KEGG pathways (Supplementary Table S1). About 65.85 % (23,004) sequences were annotated by Blast2GO with provisional names and GO categories (Table 1, Supplementary Table S1). Among them, 57.00 % was associated with molecular function, 46.42 % with biological processes, and

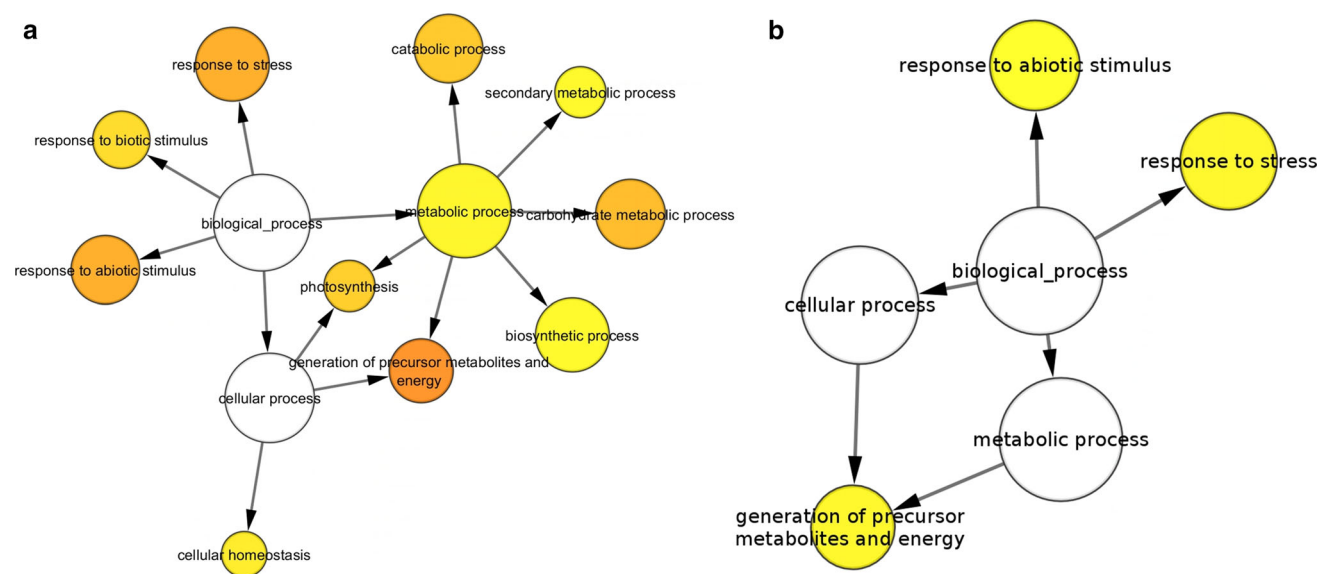
37.26 % was assigned to cellular components (Fig. 4; Supplementary Table S2). GO terms associated with primary metabolites were found, such as universal building blocks of sugars, amino acids, nucleotides, lipids, and energy sources. In addition, GO terms associated with macromolecule metabolic processes (6995), small molecule metabolic processes (5042), and protein metabolic processes (4189) were found. Analyzing the obtained transcripts, we found sequences connected to major branches of metabolism and signal transduction that are expected in complex systems (Garzon-Martinez et al. 2012). Brief summary of GO annotation is shown in Fig. 4. It is noteworthy that we found 4424 transcripts that fall into “response to stimulus” category (BP level 2), as this group includes candidate genes for pathogen resistance and many putative components of the milky sap. All unique coding sequences were used to construct reference database for further mass spectrometry analysis.

**Analysis of overrepresented GO terms in *C. majus* latex**

We utilized BiNGO to compare GO terms between *C. majus* milky sap and *C. majus* whole plant extract proteomes. The analysis comprised two steps. The first step was the comparison of *C. majus* latex proteome to *C. majus* transcriptome (Fig. 5a). The abundance of proteins involved in response to stress, response to biotic stimulus, response to abiotic stimulus, secondary metabolic process, and cellular homeostasis remains in agreement with the putative protective function of the milky sap. On the contrary, it is much harder to explain overrepresentation of photosynthesis, carbohydrate metabolic process, or generation of precursor metabolites and energy.



**Fig. 4** GO distributions for *C. majus* transcriptome. Main functional categories in the biological process (BP, level four) and molecular function (MF, level three) found in the transcriptome relevant to plant physiology



**Fig. 5** Results of comparison of *C. majus* proteomic samples using BiNGO. **a** Milky sap proteome versus transcriptome (CDS database) (significance level 0.005). **b** Milky sap proteome versus whole plant extract proteome (significance level 0.05)

Second step was the comparison of latex against whole plant extract proteomes. The results (Fig. 5b) confirmed overrepresentation of response to biotic stimulus, response to stress, and generation of precursor metabolites and energy terms.

### Analysis of *C. majus* protein composition

Previous functional assignments of *C. majus* proteins relied on homology-based information, because the sequence information of this plant was not available (Remmerie et al. 2011; Nawrot et al. 2007a, b, 2014). The novel transcriptome-based *C. majus* database allowed us to significantly improve the sensitivity of previous proteomic assessments concerning composition of *C. majus* milky sap and extracts and provide meaningful biological-based information.

For proteomic assessments using the novel database, we used two types of datasets. One of them was previously published data for proteomic content of *C. majus* whole plant extracts using an approach of protein separation by 1-D SDS-PAGE to subsequent LC-MS/MS analysis of individual gel bands (i.e., shotgun approach) (Nawrot et al. 2014). The main advantages of such approach are sample purity after SDS-PAGE improving proteome coverage of analyzed samples (Schulze and Usadel 2010; Matros et al. 2011). The other dataset comprised *C. majus* milky sap samples analyzed using the same approach. For this purpose, we prepared three independent specimens of latex (isolated from stems of 6- to 8-week-old plants) and applied shotgun approach to identify proteins (1-D gel electrophoresis, subsequent whole lane trypsin digestion

and LC-ESI-MS/MS analysis). For both types of datasets, raw proteomic data were searched against annotated *C. majus* CDS database using Mascot software. In total, 334 different putative proteins were identified for *C. majus* milky sap (for all three replicates) and 1155 for *C. majus* whole plant extract (Supplementary Tables S3, S4). The complexity of proteins in *C. majus* milky sap is hence about 3.5 times lower than in whole plant extract. The number of identified proteins from both milky sap and whole plant extract and unique CDS from transcripts were assigned to most represented KEGG pathways in *C. majus* and compared in Supplementary Table S5.

## Discussion

### *Chelidonium majus* milky sap protein composition based on comparative analysis

To demonstrate the power of the use of our new *C. majus* CDS database, we performed quantitative, comparative analysis of *C. majus* milky sap and whole plant protein samples based on emPAI obtained during Mascot search (Supplementary Table S3) to show proteins which are overrepresented in the latex, hence potentially “sap-specific”. The average of relative abundance of each protein of three latex replicates was compared with corresponding value calculated for three whole plant extract replicates (Supplementary Table S4). The resulting ratio informs about the level of overrepresentation of identified protein in the specified type of sample. The proteins observed in all sap samples with a sap/extract ratio >2 and

**Table 2** Proteins overrepresented in *C. majus* milky sap comparing to whole plant extract

No.	Transcript name	Protein name	Mean % emPAI milky sap	Mean % emPAI whole plant extract	Mean milky sap/mean whole plant extract ratio	<i>T</i> test, <i>P</i> values* <0.1
<b>Stress and defense response</b>						
1	m.37895	Mlp-like protein 28	10.32	0.00	Only in the sap	0.092
2	m.37899	Mlp-like protein 28	9.93	0.00	Only in the sap	0.047
3	m.37901	Mlp-like protein 28	2.66	0.00	Only in the sap	0.093
4	m.12634	Mlp-like protein 28	2.39	0.00	Only in the sap	0.093
5	m.12629	Mlp-like protein 34	1.73	0.03	49.91	0.094
6	m.60893	Polyphenol oxidase	1.21	0.04	32.76	0.080
7	m.12632	Mlp-like protein 28	1.05	0.01	73.34	0.094
8	m.2644	Polyphenol oxidase	0.83	0.00	Only in the sap	0.008
9	m.62963	Polyphenol chloroplastic-like	0.18	0.00	Only in the sap	0.093
10	m.18123	Mlp-like protein 28	0.12	0.00	Only in the sap	0.092
<b>Generation of precursor metabolites and energy</b>						
11	m.7838	Reticuline oxidase-like	3.73	0.05	78.19	0.073
12	m.28580	Probable caffeoyl- <i>o</i> -methyltransferase at4g26220-like	1.90	0.00	Only in the sap	0.071
13	m.2541	Anthocyanidin- <i>o</i> -glucosyltransferase-like	1.01	0.00	Only in the sap	0.061
14	m.44303	Reticuline oxidase-like	0.68	0.00	Only in the sap	0.062
15	m.44308	Reticuline oxidase-like	0.26	0.00	Only in the sap	0.099
16	m.61702	Protein usf-like	0.15	0.00	Only in the sap	0.098
<b>Stress and defense response—redox, antioxidant defense system</b>						
17	m.60799	Lactoylglutathione lyase-like	1.45	0.16	9.13	0.013
18	m.7868	Peroxidase 12-like	2.83	0.00	Only in the sap	0.061
19	m.14064	Protein in 2-1 homolog b-like	2.15	0.04	56.39	0.057
20	m.2578	Peroxidase 12-like	0.55	0.05	10.77	0.043
21	m.63454	Quinone oxidoreductase 1-like	0.51	0.05	10.95	0.079
22	uniq_03745	Isoflavone reductase homolog	0.15	0.00	Only in the sap	0.098
<b>Storage and others</b>						
23	m.54853	Beta-amylase precursor	5.83	0.17	34.17	0.038
24	m.18198	Ubiquitin-40s ribosomal protein s27a-like	3.57	0.29	12.52	0.037
25	m.60713	Lipoxygenase homology domain-containing protein 1-like	1.48	0.10	14.31	0.007
26	m.60734	Lipoxygenase homology domain-containing protein 1-like	1.10	0.11	9.64	0.030
<b>Not assigned</b>						
27	m.61326	UNCHARACTERIZED_PROTEIN	2.49	0.22	11.45	0.040
28	m.2643	UNCHARACTERIZED_PROTEIN	0.87	0.00	Only in the sap	0.092
29	m.60859	UNCHARACTERIZED_PROTEIN	0.81	0.06	12.98	0.050

Protein identification results were presented according to their relative abundance in the latex comparing to relative abundance of the same protein hits in *C. majus* whole plant extract (ratio of the average of the relative emPAI value of the latex and whole plant extract). Proteins were selected according to *t* test values ( $P < 0.1$ ), the presence of identified proteins in at least two of three sap replicates, mean sap/extract ratio >2, and mean % emPAI milky sap values >0.1. The list of all overrepresented proteins is enclosed in Supplementary Table S4

\* *t* test was used to determine if the emPAI values of selected identified proteins are significantly different between *C. majus* milky sap ( $n = 3$ ) and *C. majus* whole plant extract samples ( $n = 3$ ) with  $P < 0.1$

*t* test significance level <0.1 were considered as significantly overrepresented in *C. majus* milky sap. Overall results of the analysis are present in Supplementary

Table S4. Approximately, 29 of 334 initially identified proteins met the above-mentioned criteria and were classified as overrepresented in the milky sap (Table 2).

### Stress- and defense-related proteins overrepresented in *C. majus* milky sap

One of the overrepresented, highly abundant proteins identified only in *C. majus* milky sap, but not in *C. majus* whole plant extract, was MLP-like protein 28, which belongs to major latex protein (MLP) class (Table 2, nos. 1–4, 7, 10). It accounted for more than a quarter (26.47 %) of the protein content of the sap belonging to stress- and defense-related proteins. The second major latex protein identified was MLP-like protein 34 (Table 2, no. 5), which accounted for 1.73 % of the protein content of the milky sap. The major latex protein/ripening-related protein (MLP/RRP) subfamily is the second largest subfamily among plant proteins with 60 members, 31 of them are from *Arabidopsis thaliana* (L.) Heynh. (Radauer et al. 2008). The members of this subfamily were first described as proteins abundantly expressed in the latex of opium poppy (*Papaver somniferum* L.) (Nessler and Burnett 1992; Decker et al. 2000). The biological function of the MLP/RRP proteins is still unknown, but they have been associated with fruit and flower development and in defense or stress response (Radauer et al. 2008). Based on the modest sequence similarity, they have been characterized as members of the Bet v1 protein superfamily (Bet\_v\_I [Pfam:PF00407]) (Lytle et al. 2009). The most distinctive feature of the Bet v1 fold is a large solvent accessible hydrophobic cavity, which may function as a ligand-binding site (Radauer et al. 2008).

The other stress- and defense-related protein found in *C. majus* milky sap, which accounts for 2.22 % of the protein content of the sap, is polyphenol oxidase (PPO), which is present mainly in the latex (Table 2, nos. 6, 8–9). Two unigenes of PPO were found in *C. majus* latex: one of them is present only in the sap (Table 2, no. 8) and second is ca. 33 times more abundant in the sap than in whole plant extract (32.76-fold, Table 2, no. 6). Polyphenol oxidases or tyrosinases (PPO), also known and reported under various names (phenolase, catechol oxidase, catecholase, monophenol oxidase, *o*-diphenol oxidase, and orthophenolase) based on substrate specificity, are widely distributed in plants and fungi (Wititsuwannakul et al. 2002). The activation of PPO leads to the oxidation of phenolic compounds, consequently enhancing the resistance. PPO is also involved in the generation of reactive oxygen species (ROS). It was observed that exuding yellow *C. majus* milky sap becomes brown after exposure to air. Such browning reactions are caused by the presence of PPO family genes (Mayer 2006). Other data show that latex of different plants coagulates when exposed to air. Therefore, it is proposed that natural latex has a protective function, sealing wounds, acting as a barrier to microorganisms, and discouraging herbivory (El Moussaoui et al. 2001; Wahler et al. 2009).

The consistency of latex itself has a defensive role because the glue-like exudate seals wounds in the plant from pathogen attack and coats the mouthparts of herbivores (Hagel et al. 2008; Konno 2011; Escalante-Perez et al. 2012).

### Alkaloid and secondary metabolite biosynthetic proteins in the milky sap

*Chelidonium majus* latex is rich in a range of secondary metabolites such as alkaloids, flavonoids, or phenolic acids. *C. majus* contains alkaloids such as chelidonine, sanguinarine, chelerritrine, and berberine (Colombo and Bosisio 1996). Therefore, we identified enzymes involved in biosynthesis of these compounds: anthocyanidin-*o*-glucosyltransferase-like (Table 2, no. 13), probable caffeoyl-*o*-methyltransferase At4g26220-like (Table 2, no. 12) and reticuline oxidase-like (Table 2, nos. 11, 14–15).

Anthocyanidin-*o*-glucosyltransferase-like is an enzyme involved in anthocyanidin biosynthesis, which belongs to flavonoids. Anthocyanidins are common plant pigments. They are the sugar-free counterparts of anthocyanins forming a large group of polymethine dye (Mouradov and Spangenberg 2014).

Caffeoyl-CoA *o*-methyltransferase (EC 2.1.1.104) is an enzyme that catalyzes the reaction of conversion of *S*-adenosyl-L-methionine with caffeoyl-CoA into *S*-adenosyl-L-homocysteine and feruloyl-CoA. A large number of natural products are generated via a step involving this enzyme, which participates in phenylpropanoid biosynthesis (Boerjan et al. 2003).

Reticuline oxidase-like is the berberine bridge enzyme [(*S*)-reticuline: oxygen oxidoreductase (methylene-bridge-forming), EC 1.5.3.9], a vesicular plant enzyme that catalyzes the reaction along the biosynthetic pathway that leads to benzophenanthridine alkaloid biosynthesis. (*S*)-Reticuline is a key branch-point intermediate that can be directed into several alkaloid subtypes with different structural skeleton configurations (Ziegler et al. 2009). Cytotoxic benzophenanthridine alkaloids are accumulated in certain species of Papaveraceae and Fumariaceae in response to pathogenic attack and, therefore, function as phytoalexins (Dittrich and Kutchan 1991).

### Antioxidant and metabolic proteins in *C. majus* latex

The analysis confirmed the presence of the protein components of antioxidant defense system in the *C. majus* latex. These proteins form the first line of defense against different stress conditions and help to prevent from attack of different pathogens (Walz et al. 2002), which are highly abundant in the milky sap (Table 2, nos. 17–22). Peroxidase 12-like (Table 2, no. 18) and isoflavone reductase



homolog (Table 2, no. 22) were present only in the milky sap. The presence of class III plant peroxidase, glyoxalase, quinone reductase, and ubiquitin in *C. majus* latex was previously reported (Nawrot et al. 2007a, b).

*Chelidonium majus* latex also contains metabolic and storage proteins (Table 2, nos. 23–26). Beta-amylase precursor was relatively abundant (no. 23; 5.83 %), with 34.17-fold overrepresentation in the milky sap comparing to whole plant extract. Beta-amylase is an enzyme which hydrolyzes glucans derived from starch granules to maltose and is present in different plant organs (Fulton et al. 2008). Its presence in the milky sap could be explained by the differentiation of articulated laticifers of *C. majus*. Such gradual degeneration of the cytoplasm occurs in many species and is often accompanied by the appearance of altered plastids and characteristic starch grains (Hagel et al. 2008).

## Conclusions

Our study presents *de novo* assembly and characterization of *C. majus* transcriptome. Further comparative proteomic analysis of *C. majus* milky sap and whole plant extract samples provided new insights into milky sap protein composition. The novel transcriptome-based *C. majus* database allowed to significantly improve the sensitivity of proteomic identifications. In the present study, 334 different putative proteins were identified in *C. majus* milky sap samples comparing to only 21 in previous study (Nawrot et al. 2007a, b). Moreover, our approach enabled identification of the major latex protein (MLP) 28, which could not be detected without species-specific database. The quantitative analysis confirmed that *C. majus* latex is rich in proteins connected with response to stress conditions and generation of precursor metabolites and energy. We also identified polyphenol oxidase (PPO), several enzymes involved in biosynthesis of natural products and a range of abundant antioxidant proteins. These findings support the importance of *C. majus* latex for plant defense against pathogens and herbivores.

The obtained *C. majus* annotated CDS database will serve as a valuable dataset for further studies of *C. majus* proteomes and as comparative material for other plant species.

The data sets supporting the results of this article are available at the NCBI Sequence Read Archive (SRA) and are available under SRA Accession Number SRR1998045 (related BioProject PRJNA264791—*Chelidonium majus* transcriptome, related BioSample SAMN03142649).

**Author contribution statement** RN performed the study, collected plant material, performed proteomic analyses and prepared the manuscript. JB performed gene annotation,

bioinformatics and comparative analyses. RL participated in design and coordination of the study, collected plant material. LA performed sequencing and initial assembly. HPM supervised the work, participated in its design and coordination and corrected the manuscript. All authors read and approved the final manuscript.

**Acknowledgments** The research was supported by the National Science Centre Grant No. 2011/03/B/NZ9/01335 to R. Nawrot (Adam Mickiewicz University in Poznań, Faculty of Biology, Poland) and Grant No. N N405 677740 to B. Kędzia (Institute of Natural Fibers and Medicinal Plants, Poznań, Poland). Robert Nawrot's stay in IPK Gatersleben was supported by "AMU: A Unique Graduate = Possibilities" scholarship for young post-doc researchers funded by EU—European Social Fund No. POKL.04.01.01-00-019/10. Special acknowledgments to VitaInSilica Company, Poznań, Poland, for help with *de novo* *C. majus* transcriptome assembly and annotation. H-P Mock gratefully acknowledges funding by the BMBF (EULAFUEL, FKZ 0315699). Special acknowledgments for J. Dębski from the Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, for help with proteomic data analysis. The equipment used was sponsored in part by the Centre for Preclinical Research and Technology (CePT), a project co-sponsored by European Regional Development Fund and Innovative Economy, The National Cohesion Strategy of Poland.

## Compliance with ethical standards

**Conflict of interest** The authors declare no conflicts of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54:519–546
- Colombo ML, Bosisio E (1996) Pharmacological activities of *Chelidonium majus* L. (Papaveraceae). *Pharmacol Res* 33:127–134
- Decker G, Wanner G, Zenk MH, Lottspeich F (2000) Characterization of proteins in latex of the opium poppy (*Papaver somniferum*) using two-dimensional gel electrophoresis and microsequencing. *Electrophoresis* 21:3500–3516
- Dittrich H, Kutchan TM (1991) Molecular-cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. *P Natl Acad Sci USA* 88:9969–9973
- El Moussaoui A, Nijs M, Paul C, Wintjens R, Vincentelli J, Azarkan M, Looze Y (2001) Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism. *Cell Mol Life Sci* 58:556–570
- Escalante-Perez M, Jaborsky M, Lautner S, Fromm J, Muller T, Dittrich M, Kunert M, Boland W, Hedrich R, Ache P (2012) Poplar extrafloral nectaries: two types, two strategies of indirect defenses against herbivores. *Plant Physiol* 159:1176–1191

- Fu CH, Chen YW, Hsiao YY, Pan ZJ, Liu ZJ, Huang YM, Tsai WC, Chen HH (2011) OrchidBase: a collection of sequences of the transcriptome derived from orchids. *Plant Cell Physiol* 52:238–243
- Fugate KK, Fajardo D, Schlautman B, Ferrareze JP, Bolton MD, Campbell LG, Wiesman E, Zalapa J (2014) Generation and characterization of a sugarbeet transcriptome and transcript-based SSR markers. *Plant Genome*. doi:10.3835/plantgenome2013.11.0038
- Fulton DC, Stettler M, Mettler T, Vaughan CK, Li J, Francisco P, Gil D, Reinhold H, Eicke S, Messerli G, Dorken G, Halliday K, Smith AM, Smith SM, Zeeman SC (2008) beta-AMYLASE4, a noncatalytic protein required for starch breakdown, acts upstream of three active beta-amylases in *Arabidopsis* chloroplasts. *Plant Cell* 20:1040–1058
- Garzon-Martinez GA, Zhu ZI, Landsman D, Barrero LS, Marino-Ramirez L (2012) The *Physalis peruviana* leaf transcriptome: assembly, annotation and gene model prediction. *BMC Genom* 13:151. doi:10.1186/1471-2164-13-151
- Gong L, Zhang HX, Gan XY, Zhang L, Chen YC, Nie FJ, Shi L, Li M, Guo ZQ, Zhang GH, Song YX (2015) Transcriptome profiling of the potato (*Solanum tuberosum* L.) plant under drought stress and water-stimulus conditions. *PLoS One* 10(5):e0128041. doi:10.1371/journal.pone.0128041
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Hagel JM, Yeung EC, Facchini PJ (2008) Got milk? The secret life of laticifers. *Trends Plant Sci* 13:631–639
- Hamilton JP, Buell CR (2012) Advances in plant genome sequencing. *Plant J* 70:177–190
- Hao DC, Ge GB, Xiao PG, Zhang YY, Yang L (2011) The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PLoS One* 6(6):e21220. doi:10.1371/journal.pone.0021220
- Hiller KO, Ghorbani M, Schilcher H (1998) Antispasmodic and relaxant activity of chelidonine, protopine, coptisine, and *Chelidonium majus* extracts on isolated guinea-pig ileum. *Planta Med* 64:758–760
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteom* 4:1265–1272
- Jayakodi M, Lee SC, Lee YS, Park HS, Kim NH, Jang W, Lee HO, Joh HJ, Yang TJ (2015) Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biol* 15:138. doi:10.1186/s12870-015-0527-0
- Konno K (2011) Plant latex and other exudates as plant defense systems: roles of various defense chemicals and proteins contained therein. *Phytochemistry* 72:1510–1530
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685
- Lytle BL, Song J, de la Cruz NB, Peterson FC, Johnson KA, Bingman CA, Phillips GN Jr, Volkman BF (2009) Structures of two *Arabidopsis thaliana* major latex proteins represent novel helix-grip folds. *Proteins* 76:237–243
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449
- Matros A, Kaspar S, Witzel K, Mock HP (2011) Recent progress in liquid chromatography-based separation and label-free quantitative plant proteomics. *Phytochemistry* 72:963–974
- Mayer AM (2006) Polyphenol oxidases in plants and fungi: going places? A review. *Phytochemistry* 67:2318–2331
- Mouradov A, Spangenberg G (2014) Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci* 5:620. doi:10.3389/fpls.2014.00620
- Nawrot R, Kalinowski A, Gozdzicka-Jozefiak A (2007a) Proteomic analysis of *Chelidonium majus* milky sap using two-dimensional gel electrophoresis and tandem mass spectrometry. *Phytochemistry* 68:1612–1622
- Nawrot R, Lesniewicz K, Pienkowska J, Gozdzicka-Jozefiak A (2007b) A novel extracellular peroxidase and nucleases from a milky sap of *Chelidonium majus*. *Fitoterapia* 78:496–501
- Nawrot R, Zauber H, Schulze WX (2014) Global proteomic analysis of *Chelidonium majus* and *Corydalis cava* (Papaveraceae) extracts revealed similar defense-related protein compositions. *Fitoterapia* 94:77–87
- Nessler CL, Burnett RJ (1992) Organization of the major latex protein gene family in opium poppy. *Plant Mol Biol* 20:749–752
- Neuhoff V, Arold N, Taube D, Ehrhardt W (1988) Improved staining of proteins in polyacrylamide gels including isoelectric-focusing gels with clear background at nanogram sensitivity using coomassie brilliant blue G-250 and R-250. *Electrophoresis* 9:255–262
- Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567
- Radauer C, Lackner P, Breiteneder H (2008) The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol* 8:286. doi:10.1186/1471-2148-8-286
- Remmerie N, De Vijlder T, Laukens K, Dang TH, Lemièrre F, Mertens I, Valkenburg D, Blust R, Witters E (2011) Next generation functional proteomics in non-model plants: a survey on techniques and applications for the analysis of protein complexes and post-translational modifications. *Phytochemistry* 72:1192–1218
- Saito R, Smoot ME, Ono K, Ruschinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9:1069–1076
- Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. *Annu Rev Plant Biol* 61:491–516
- Wahler D, Gronover CS, Richter C, Foucu F, Twyman RM, Moerschbacher BM, Fischer R, Muth J, Pruber D (2009) Polyphenoloxidase silencing affects latex coagulation in *Taraxacum* species. *Plant Physiol* 151:334–346
- Walz C, Juenger M, Schad M, Kehr J (2002) Evidence for the presence and activity of a complete antioxidant defence system in mature sieve tubes. *Plant J* 31:189–197
- Wititsuwannakul D, Chareonthiphakorn N, Pace M, Wititsuwannakul R (2002) Polyphenol oxidases from latex of *Hevea brasiliensis*: purification and characterization. *Phytochemistry* 61:115–121
- Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, Zhu YJ, Ji AJ, Zhang B, Hu SN, Au KF, Song JY, Chen SL (2015) Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J* 82:951–961
- Zeng JG, Liu YS, Liu W, Liu XB, Liu FQ, Huang P, Zhu PC, Chen JJ, Shi MM, Guo F, Cheng P, Zeng J, Liao YF, Gong J, Zhang HM, Wang DP, Guo AY, Xiong XY (2013) Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata* and *Macleaya microcarpa*. *PLoS One* 8(1):e53409. doi:10.1371/journal.pone.0053409
- Ziegler J, Facchini PJ, Geissler R, Schmidt J, Ammer C, Kramell R, Voigtlander S, Gesell A, Pienkny S, Brandt W (2009) Evolution of morphine biosynthesis in opium poppy. *Phytochemistry* 70:1696–1707