JOURNAL OF NATURAL PRODUCTS

Article

pubs.acs.org/jnp

The Combined Use of Proteomics and Transcriptomics Reveals a Complex Secondary Metabolite Network in *Peperomia obtusifolia*

Andrea N. L. Batista,*[,†] José Roberto A. dos Santos-Pinto,[‡] João M. Batista, Jr.,[§] Tatiana M. Souza-Moreira,[⊥] Mariana M. Santoni,[⊥] Cleslei F. Zanelli,[⊥] Massuo J. Kato,[∥] Silvia N. López,[▽] Mario S. Palma,[‡] and Maysa Furlan*[,†]

[†]Instituto de Química, Universidade Estadual Paulista (Unesp), Araraquara, SP 14800-060, Brazil

[‡]Instituto de Biociências, Universidade Estadual Paulista (Unesp), Rio Claro, SP 13506-900, Brazil

[§]Departamento de Química, Universidade Federal de São Carlos−UFSCar, São Carlos, SP 13565-905, Brazil

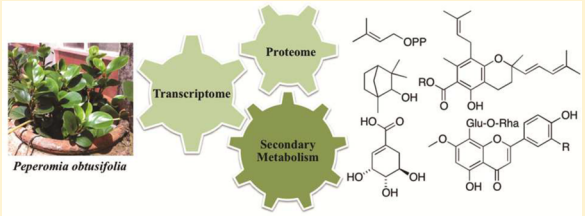
[⊥]Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (Unesp), Araraquara, SP 14801-902, Brazil

[∥]Instituto de Química, Universidade de São Paulo−USP, São Paulo, SP 05508-000, Brazil

[▽]CONICET, Farmacognosia, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Santa Fe S2002LRK, Argentina

Supporting Information

ABSTRACT: *Peperomia obtusifolia*, an ornamental plant from the Piperaceae family, accumulates a series of secondary metabolites with interesting biological properties. From a biosynthesis standpoint, this species produces several benzopyrans derived from orsellinic acid, which is a polyketide typically found in fungi. Additionally, the chiral benzopyrans were reported as racemic and/or as diastereomeric mixtures, which raises questions about the level of enzymatic control in the cyclization step for the formation of the 3,4-dihydro-2*H*-pyran moiety. Therefore, this article describes the use of shotgun proteomic and transcriptome studies as well as phytochemical profiling for the characterization of the main biosynthesis pathways active in *P. obtusifolia*. This combined approach resulted in the identification of a series of proteins involved in its secondary metabolism, including tocopherol cyclase and prenyltransferases. The activity of these enzymes was supported by the phytochemical profiling performed in different organs of *P. obtusifolia*. However, the polyketide synthases possibly involved in the production of orsellinic acid could not be identified, suggesting that orsellinic acid may be produced by endophytes intimately associated with the plant.

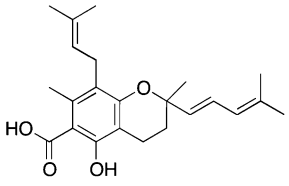


Peperomia obtusifolia is a Crassulacean acid metabolism plant (CAM) that belongs to the Piperaceae family and grows in regions from Mexico to the northern parts of South America.[1] It is popularly known as *paragua* or *baby rubber* plant. Despite its predominant ornamental usage, some communities in Central America use its leaves, stems, and fruits to treat insect and snake bites and also as a skin cleanser.[2] Phytochemical studies of *P. obtusifolia* have led to the isolation of bioactive secondary metabolites such as prenylated chromans, lignans, amides, flavonoids, and other phenolic derivatives.[1,3,4] A survey of biologically active metabolites from Piperaceae species revealed that the *Peperomia* genus is the second best source of bioactive compounds within this family, contributing as high as 15% of the total bioactive compounds.[5] This number could be higher given that only a few studies have been carried out on this genus, mainly due to its predominantly ornamental use. The same study also showed that prenylated benzopyrans are among the most potent biologically active compounds found in *Peperomia* species, with cytotoxic and antiprotozoal activities being the most frequently reported.[5] Interestingly, this species appears to produce orsellinic acid-derived benzopyrans (**1**−**4**), with some isolated as racemic and/or diastereomeric mixtures.[4,6,7] Orsellinic acid is a precursor commonly found in fungi,[8] and its isolation from plant extracts may suggest horizontal gene transfers from endophytic fungi to the host plant.[9] It may also indicate that a given metabolite has been actually produced by endophytes intimately associated with the plant.[10] Additionally, the absence of enantiomerically pure/enriched benzopyrans in *P. obtusifolia* calls into question the level of enzymatic control involved in the intramolecular cyclization that yields the 3,4-dihydro-2*H*-pyran unit.[11]

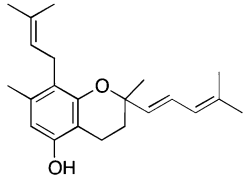
In order to improve the understanding of the main processes involved in the secondary metabolism of *P. obtusifolia*, it is crucial to identify the majority of the proteins expressed in this species. In this context, it is particularly important to search for proteins related to the biosynthesis of orsellinic acid as well as to the cyclization of the prenylated orsellinic acid precursor. To that end, the shotgun proteomic approach is one of the most

Received: September 8, 2016

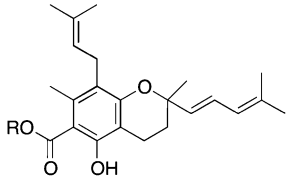
DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX



1



2



R= borneol (3) or fenchol (4)

powerful methodologies to analyze complex protein mixtures in a single experiment.[12] However, for organisms whose genome has not been sequenced, as in the case of *P. obtusifolia*, the identification of proteins is performed through comparisons against protein databanks of cross-species.[13] As an alternative source of information we have also carried out a comprehensive transcriptome study focused on genes encoding secondary metabolism proteins. RNA-Seq (high-performance sequencing) is an approach to transcriptome profiling that has been particularly useful in nonmodel species where genetic and genomic resources are limited and little or no sequence data are available.[14]

Herein we describe the combination of shotgun proteomics of the soluble and microsomal fractions, transcriptome studies, and phytochemical profiling of different plant compartments aimed at identifying the main biosynthesis pathways active in *P. obtusifolia*.

RESULTS AND DISCUSSION

Shotgun Proteomics. The shotgun proteomic analyses of the soluble fractions of the different organs from *P. obtusifolia* led to the identification of 325 proteins from the leaves, 290 from the stems, and 311 from the roots. A complete list of all the proteins identified is provided in the Supporting Information (Tables S1–S3). Notably, only 18 proteins were found to be common among the organs, with the leaves and stems exhibiting the highest similarity (Figure S1A, Supporting Information). To identify functional categories, the proteins identified were assigned to gene ontology (GO) terms and mapped at varying levels of detail. As this work aims to identify sets of transformations that explain the presence of some groups of secondary metabolites rather than investigating individual transformation mechanisms or enzymes, only the category "biological process" was analyzed. Figure 1 shows the main biological processes (level 3 of detail) identified for each organ and the number of proteins involved in each process. In this type of analysis the same proteins may be assigned to more than one category (Tables S4–S6, Supporting Information). Overall, the distribution of the biological processes is similar in the three organs studied. As expected, proteins involved in cellular and metabolic processes were dominant in all the organs. Proteins involved in response to stimulus also appear as an abundant subcategory, consisting of 117 proteins in the leaves, 116 in the stems, and 118 in the roots. These latter proteins perform a variety of functions, playing an important

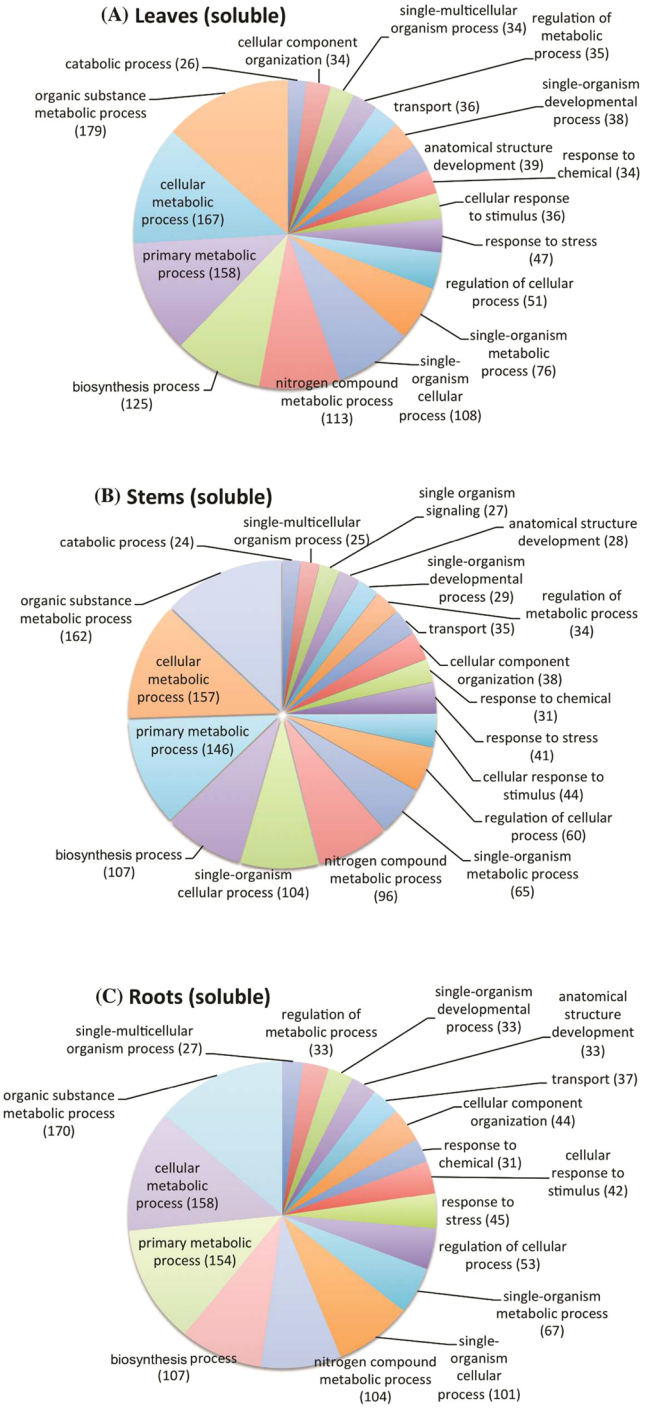


Figure 1. Functional category distribution based on "biological process" of the identified proteins using shotgun proteomic analysis. Number of proteins expressed in soluble fractions of leaves (A), stems (B), and roots (C).

role in plant defense against pathogenic restrictions and general adaptation to stress.[15] Around 40 transport proteins were found in each organ.

A number of soluble proteins related to the biosynthesis of secondary metabolites were also identified in the leaves, stems, and roots (Table 1). These proteins were grouped within the subcategory biosynthesis process (Tables S4–S6, Supporting Information). The organ with the largest number of identified

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

Table 1. Proteins Involved in the Biosynthesis of Secondary Metabolites Identified in the Soluble Fractions of the Leaves, Stems, and Roots of *P. obtusifolia* Using Shotgun Proteomic Analysis

protein	soluble fractions			biological process
	leaves	stems	roots	
acetyl-CoA carboxylase 1			×	malonyl-CoA biosynthesis
acetyl-CoA carboxylase 2		×		malonyl-CoA biosynthesis
acetyl-CoA carboxylase carboxyl transferase subunit alpha			×	malonyl-CoA biosynthesis
acetyl-CoA carboxylase carboxyl transferase subunit beta	×	×		malonyl-CoA biosynthesis
3-hydroxy-3-methylglutaryl-coenzyme A reductase 2	×			isoprene biosynthesis
4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase			×	isoprene biosynthesis
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase			×	isoprene biosynthesis
geranylgeranyl pyrophosphate synthase		×		isoprenoid biosynthesis
squalene synthase			×	isoprenoid biosynthesis
limonene synthase		×	×	monoterpene biosynthesis
(−)-camphene/tricyclene synthase	×			monoterpene biosynthesis
1,8-cineole synthase 1		×		monoterpene biosynthesis
(+)-neomenthol dehydrogenase			×	monoterpene biosynthesis
α-copaene synthase	×			sesquiterpene biosynthesis
inactive sesquithujene synthase B		×		sesquiterpene biosynthesis
(*S*)-*β*-macrocarpene synthase		×		sesquiterpene biosynthesis
β-caryophyllene synthase			×	sesquiterpene biosynthesis
ent-kaurene synthase-like 5	×			diterpene biosynthesis
casbene synthase			×	diterpene biosynthesis
gibberellin 20 oxidase 3	×			diterpene biosynthesis
β-amyrin 28-oxidase	×			triterpene biosynthesis
cycloartenol synthase		×	×	steroid biosynthesis
chorismate mutase 2		×	×	chorismate biosynthesis
phenylalanine ammonia-lyase (PAL)		×		phenylpropanoid metabolism
probable 4-coumarate-CoA ligase		×	×	phenylpropanoid metabolism
trans-anol *O*-methyltransferase 1		×		phenylpropanoid metabolism
polyketide synthase 5			×	flavonoid biosynthesis
dihydroflavonol-4-reductase			×	flavonoid biosynthesis
bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase			×	flavonoid biosynthesis
leucoanthocyanidin dioxygenase			×	anthocyanin biosynthesis
anthocyanidin 3-*O*-glucosyl transferase 1			×	anthocyanin biosynthesis
anthocyanidin 3-*O*-glucoside 5-*O*-glucosyltransferase			×	anthocyanin biosynthesis
malonyl CoA:anthocyanin 5-*O*-glucoside-6‴-*O*-malonyltransferase	×			anthocyanin biosynthesis
cytochrome P450 71A26	×			secondary metabolite biosynthesis
cytochrome P450 71B17	×	×		secondary metabolite biosynthesis
cytochrome P450 71B19	×	×		secondary metabolite biosynthesis
cytochrome P450 71B28		×		secondary metabolite biosynthesis
cytochrome P450 71B34		×		secondary metabolite biosynthesis
cytochrome P450 71B37	×			secondary metabolite biosynthesis
cytochrome P450 76C3			×	secondary metabolite biosynthesis
cytochrome P450 71D8		×		secondary metabolite biosynthesis
cytochrome P450 71D9	×		×	secondary metabolite biosynthesis

proteins related to the biosynthesis of secondary metabolites was the roots (20), followed by the stems (17) and leaves (13).

However, important enzymes expected to participate both in the biosynthesis of orsellinic acid (orsellinic acid synthase; polyketide synthases) and in the intramolecular cyclization of a prenylated/geranylated orsellinic acid precursor yielding the chiral 3,4-dihydro-2*H*-pyran moiety (tocopherol cyclase-like enzymes) were not identified. Therefore, to complete the proteome analysis, shotgun proteomics was also applied to the microsomal fractions of *P. obtusifolia*, particularly because many cyclases and prenyltransferases are commonly associated with the membranes.[16,17]

The shotgun proteomic analyses of the microsomal fractions from different organs of *P. obtusifolia* led to the identification of 301 proteins expressed in the leaves, 276 expressed in the stems, and 303 expressed in the roots. Again, the greatest similarity occurs between leaves and stems (Figure S1B, Supporting Information). Only 27 proteins were common to soluble and microsomal fractions from the leaves, an equal number of proteins were common to these fractions from the stems, and only 18 proteins were common among the two different fractions from the roots. Despite the low similarity in protein identity between the soluble and microsomal fractions, their related biological processes were quite similar, mostly on stems (Figure 2). As expected, cellular processes and metabolic processes were the main subcategories of proteins identified in the three organs studied. Other categories that showed a considerable number of proteins, mainly in the leaves, were responses to stimulus. As for transport proteins, they occur in largest quantity in the roots, followed by leaves and stems. The amount of secondary metabolism proteins in the microsomal fractions was considerably higher than that in the soluble fractions (41 in the leaves, 39 in the stems, and 54 in the roots compared to an overall 13−20 for the soluble fractions). Roots were the most representative organ once again. All enzymes identified in these fractions that were involved in the biosynthesis of secondary metabolites are summarized in Table 2. A complete list of the proteins identified from the microsomal fraction and the subcategories of biological processes to which they are related can be found in the Supporting Information (Tables S7−S9 and S10−S12, respectively).

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

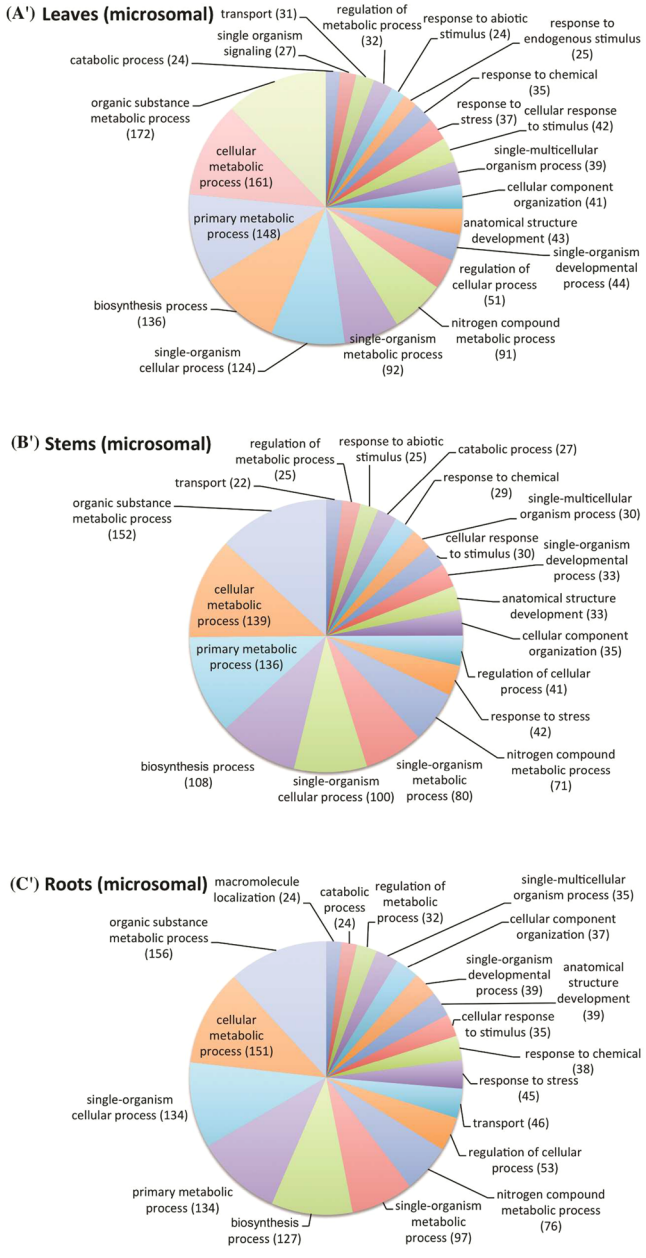


Figure 2. Functional category distribution based on "biological process" of the identified proteins using shotgun proteomic analysis. Number of proteins expressed in microsomal fractions of leaves (A′), stems (B′), and roots (C′).

The combination of shotgun proteomic data of soluble and microsomal fractions led to the characterization of many biosynthesis pathways active in *P. obtusifolia*. Enzymes related to glycolysis and to the formation of shikimic acid, acetyl-CoA, and malonyl-CoA, as well as isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), were identified (Figure S2, Supporting Information).

Regarding secondary metabolites, the biosynthesis of terpenes (Figure S3, Supporting Information) was the pathway with the largest number of enzymes identified. Some enzymes were involved in the early stages of terpene formation, such as farnesyl pyrophosphate synthase (leaves), squalene synthase (leaves, stems, and roots), and squalene epoxidase (roots). Eleven enzymes related to the biosynthesis of monoterpenes were identified including (−)-*endo*-fenchol synthase (roots), which may be involved in the biosynthesis of the monoterpene moiety of compound **4**. A number of enzymes related to the biosynthesis of sesquiterpenes and diterpenes were found, including enzymes involved in the biosynthesis of gibberellins. Enzymes involved in nearly all the steps of the biosynthesis of jasmonic acid were also identified in the roots. Jasmonate and gibberellic acid are important plant hormones that mediate defense and growth, respectively.[18] In addition to squalene synthase and squalene epoxidase, various enzymes related to the biosynthesis of steroids and triterpenoids were identified (Tables 1 and 2).

With this analysis, it was also possible to verify the occurrence of several enzymes involved in the biosynthesis of phenylpropanoids (Figure S4, Supporting Information), including shikimate kinase (roots) and phenylalanine ammonia-lyase (leaves, stems, and roots). Other important enzymes of this route catalyzing the formation of flavonoids and lignans were identified, such as cinnamyl alcohol dehydrogenase (leaves, stems, and roots) and caffeic acid methyltransferase (roots) (Tables 1 and 2). These results corroborate the literature data on the occurrence of phenylpropanoid derivatives such as flavonoids and lignans in *P. obtusifolia*.[3,19]

Of particular interest, several enzymes putatively involved in the biosynthesis of the flavonoids 4′-*O*-methylisoswertisin-2″-*α*-L-rhamnoside and isoswertisin-2″-*α*-L-rhamnoside, whose occurrence in this species has already been reported,[19] were identified (Figure S5, Supporting Information). These enzymes were polyketide synthase 5 (roots), chalcone synthase (leaves, stems, and roots), chalcone-flavanone isomerase A (leaves), and narigenin-7-*O*-methyltransferase (leaves). Furthermore, the biosynthesis pathways for the production of isoflavonoids and anthocyanins were also identified (Figure S5, Supporting Information).

Interestingly, no enzymes directly related to the biosynthesis of the chromans, such as chroman cyclases and/or orsellinic acid synthase, were identified by shotgun proteomics. However, as mentioned before, the enzyme (−)-*endo*-fenchol synthase, which may be involved in the biosynthesis of the monoterpene moiety of compound **4**, was found in the roots (Figure 3).

Transcriptome Studies. Following proteome analysis, *P. obtusifolia* leaf transcriptomic data were generated using a next-generation sequencing technology. The filtered reads were assembled *de novo* into 121 017 transcripts. A total of 81 831 transcripts (68%) were between 200 and 1000 bases, followed by 24 900 transcripts (21%) ranging from 1000 to 2000 bases. The transcript assembly details are summarized in Table 3. These results were comparable to the recently published plant transcriptomes from *Piper nigrum*,[20] *Reaumuria* soongorica,[21] and *Haloxylon ammodendron*,[22] which were assembled using the same type of sequencing platform and suggest that the sequencing output was of sufficient quality for further analysis.

Primarily, the assembled transcripts were annotated against the custom plant protein database and 66 591 transcripts (55%) had significant hits (E-value $< 1 \times 10^{-5}$). To gain further insight into secondary metabolism, the annotated transcripts were analyzed by searching them against the KEGG database. Out of the 66 591 annotated transcripts, 19 304 (28.98%) were identified with a KEGG number (*k*-number) corresponding to 3287 distinct genes. Among them, 856 (26.04%) were related to metabolic pathways and 395 (12.02%) to secondary metabolism. However, only the transcripts with FPKM ≥ 1 were considered, resulting in 286 genes encoding enzymes involved in secondary metabolism (Table S13, Supporting

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

Table 2. Proteins Involved in the Biosynthesis of Secondary Metabolites Identified in the Microsomal Fractions of the Leaves, Stems, and Roots of *P. obtusifolia* Using Shotgun Proteomic Analysis

protein	microsomal fractions			biological process
	leaves	stems	roots	
acetyl-CoA carboxylase 1			×	malonyl-CoA biosynthesis
acetyl-CoA carboxylase 2	×	×	×	malonyl-CoA biosynthesis
acetyl-CoA carboxylase carboxyl transferase subunit beta	×		×	malonyl-CoA biosynthesis
polyketide synthase type III	×			malonyl-CoA biosynthesis
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase			×	isoprene biosynthesis
3-hydroxy-3-methylglutaryl-coenzyme A reductase 1			×	isoprene biosynthesis
farnesyl pyrophosphate synthase	×			isoprenoid biosynthesis
squalene synthase	×	×	×	isoprenoid biosynthesis
squalene monooxygenase 1,2			×	isoprenoid biosynthesis
squalene epoxidase 3			×	isoprenoid biosynthesis
terpenoid synthase 1	×			terpene biosynthesis
terpenoid synthase 15			×	terpene biosynthesis
probable terpene synthase 4	×			terpene biosynthesis
(−)-α-pinene synthase	×	×		monoterpene biosynthesis
1,8-cineole synthase	×			monoterpene biosynthesis
myrcene synthase	×		×	monoterpene biosynthesis
geraniol 8-hydroxylase		×		monoterpene biosynthesis
geraniol dehydrogenase 1		×		monoterpene biosynthesis
(*E*)-β-ocimene synthase		×		monoterpene biosynthesis
(−)-camphene/tricyclene synthase		×		monoterpene biosynthesis
terpinolene synthase		×		monoterpene biosynthesis
(*R*)-linalool synthase QH1			×	monoterpene biosynthesis
limonene synthase			×	monoterpene biosynthesis
camphene synthase			×	monoterpene biosynthesis
α-terpineol synthase			×	monoterpene biosynthesis
(−)-*endo*-fenchol synthase			×	monoterpene biosynthesis
α-bisabolol synthase	×	×		sesquiterpene biosynthesis
viridiflorene synthase	×			sesquiterpene biosynthesis
bicyclogermacrene synthase	×			sesquiterpene biosynthesis
γ-cadinene synthase	×			sesquiterpene biosynthesis
(*E*)-γ-bisabolene synthase		×		sesquiterpene biosynthesis
(Z)-γ-bisabolene synthase 2	×			sesquiterpene biosynthesis
(*E*)-β-farnesene synthase		×		sesquiterpene biosynthesis
germacrene-A synthase		×		sesquiterpene biosynthesis
germacrene A synthase short form			×	sesquiterpene biosynthesis
germacrene A oxidase			×	sesquiterpene biosynthesis
germacrene-D synthase			×	sesquiterpene biosynthesis
germacrene-D synthase 2			×	sesquiterpene biosynthesis
β-cubebene synthase			×	sesquiterpene biosynthesis
(*S*)-β-macrocarpene synthase			×	sesquiterpene biosynthesis
artemisinic aldehyde delta(11(13)) reductase			×	sesquiterpene biosynthesis
epi-cedrol synthase			×	sesquiterpene biosynthesis
bifunctional abietadiene synthase		×		diterpene biosynthesis
bifunctional *cis*-abienol synthase		×	×	diterpene biosynthesis
ent-kaurene synthase-like 5		×		diterpene biosynthesis
ent-kaurenoic acid oxidase 1		×		diterpene biosynthesis
gibberellin 2-β-dioxygenase 2			×	diterpene biosynthesis
gibberellin 2-β-dioxygenase 3		×		diterpene biosynthesis
gibberellin 20 oxidase 1		×		diterpene biosynthesis
gibberellin 20 oxidase 1-A			×	diterpene biosynthesis
gibberellin 20 oxidase 1-B	×			diterpene biosynthesis
gibberellin 20 oxidase 2		×	×	diterpene biosynthesis
gibberellin 20 oxidase 3	×			diterpene biosynthesis
gibberellin 20 oxidase 4			×	diterpene biosynthesis
taraxerol synthase	×			triterpene biosynthesis
lupeol synthase	×	×		triterpene biosynthesis
camelliol C synthase	×		×	triterpene biosynthesis
β-amyrin synthase	×	×	×	triterpene biosynthesis

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

Table 2. continued

protein	microsomal fractions			biological process
	leaves	stems	roots	
β-amyrin 28-oxidase	×	×		triterpene biosynthesis
amyrin synthase LUP2	×			triterpene biosynthesis
thalianol synthase	×	×	×	triterpene biosynthesis
dammaradiene synthase		×		triterpene biosynthesis
arabidiol synthase		×		triterpene biosynthesis
cycloartenol synthase		×	×	steroid biosynthesis
cycloartenol-*C*-24-methyltransferase 1	×			steroid biosynthesis
methylsterol monooxygenase 1−3	×			steroid biosynthesis
3-*β*-Hydroxysteroid-dehydrogenase/decarboxylase		×		steroid biosynthesis
chorismate synthase 2	×			chorismate biosynthesis
shikimate kinase 3			×	chorismate biosynthesis
phenylalanine ammonia-lyase	×	×		phenylpropanoid metabolism
phenylalanine ammonia-lyase 1			×	phenylpropanoid metabolism
phenylalanine ammonia-lyase 3	×			phenylpropanoid metabolism
phenylalanine ammonia-lyase G2B		×		phenylpropanoid metabolism
cinnamyl alcohol dehydrogenase 1	×			phenylpropanoid metabolism
cinnamyl alcohol dehydrogenase 2		×		phenylpropanoid metabolism
cinnamyl alcohol dehydrogenase 5			×	phenylpropanoid metabolism
cinnamyl alcohol dehydrogenase 9			×	phenylpropanoid metabolism
trans-cinnamate 4-monooxygenase		×	×	phenylpropanoid metabolism
4-coumarate-CoA ligase 1			×	phenylpropanoid metabolism
4-coumarate-CoA ligase-like 2		×	×	phenylpropanoid metabolism
4-coumarate–CoA ligase 3			×	phenylpropanoid metabolism
caffeic acid 3-*O*-methyltransferase			×	phenylpropanoid metabolism
chalcone synthase		×	×	flavonoid biosynthesis
chalcone synthase A (fragment)		×		flavonoid biosynthesis
chalcone synthase 7	×	×		flavonoid biosynthesis
chalcone-flavanone isomerase A	×			flavonoid biosynthesis
naringenin 7-*O*-methyltransferase	×			flavonoid biosynthesis
flavonol synthase/flavanone 3-hydroxylase			×	flavonoid biosynthesis
flavonoid 3′,5′-hydroxylase	×		×	flavonoid biosynthesis
flavonoid 3′,5′-hydroxylase 2			×	flavonoid biosynthesis
dihydroflavonol-4-reductase	×			flavonoid biosynthesis
2-hydroxyisoflavanone synthase			×	isoflavonoid biosynthesis
isoflavone reductase homologue 2			×	isoflavonoid biosynthesis
isoflavone 4′-*O*-methyltransferase	×			isoflavonoid biosynthesis
flavonoid 3-*O*-glucosyltransferase		×		flavonoid biosynthesis
UDP-glycosyltransferase 71B8		×		flavonoid biosynthesis
UDP-glycosyltransferase 76F2	×	×		flavonoid biosynthesis
UDP-glycosyltransferase 87A2		×		flavonoid biosynthesis
leucoanthocyanidin reductase			×	proanthocyanidin biosynthesis
catechol oxidase B			×	anthocyanin biosynthesis
anthocyanidin 3-*O*-glucosyltransferase 5	×			anthocyanin biosynthesis
anthocyanidin 3-*O*-glucosyltransferase 6			×	anthocyanin biosynthesis
anthocyanidin 3-*O*-glucosyltransferase 7			×	anthocyanin biosynthesis
cyanidin-3-*O*-glucoside 2-*O*-glucuronosyltransferase			×	anthocyanin biosynthesis
malonyl-coenzyme A:anthocyanin 3-*O*-glucoside-6″-*O*-malonyltransferase	×			anthocyanin biosynthesis
coumaroyl-CoA:anthocyanidin 3-*O*-glucoside-6″-*O*-coumaroyltransferase 1	×			anthocyanin biosynthesis
anthocyanidin 3-*O*-glucoside 5-*O*-glucosyltransferase 1			×	anthocyanin biosynthesis
pelargonidin 3-*O*-(6-caffeoylglucoside) 5-*O*-(6-*O*-malonylglucoside) 4‴-malonyltransferase	×			anthocyanin biosynthesis
anthocyanin 5-aromatic acyltransferase			×	anthocyanin biosynthesis

Information). These genes were mapped into 40 KEGG metabolic pathways related to secondary metabolism that includes carbohydrate and lipid metabolism and nucleotide and amino acid metabolism, among others. As a result, several metabolic pathways were identified with terpenoid biosynthesis representing the largest group.

The results of the transcriptome analysis of the leaves from *P. obtusifolia* supported the data obtained from shotgun proteomics. As shown in the KEGG maps, the transcriptome analysis permitted confirmation of the presence of several enzymes related to glycolysis and formation of biosynthesis "building blocks" such as shikimic acid, acetyl and malonyl-

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

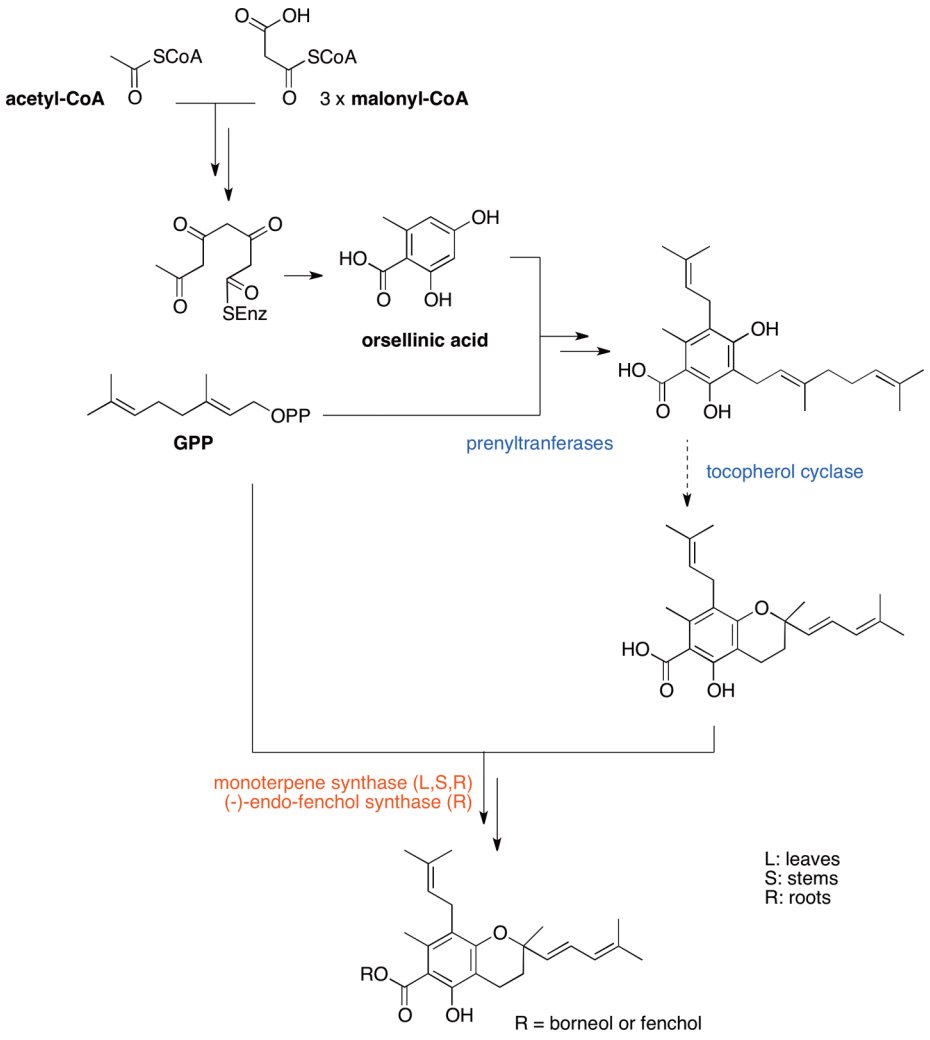


Figure 3. Schematic representation of the biosynthesis of chromans in *P. obtusifolia* including the main proteins identified. Red font: Enzymes identified by shotgun proteomics. Blue font: Enzymes identified from transcriptomic studies.

Table 3. RNA Sequencing Summary of *Peperomia obtusifolia* Transcriptome

description	*Peperomia obtusifolia* (leaves)
total number of raw reads[a]	36 330 714
number of paired-end reads after trimming[a]	31 888 176
% bases ≥ Q30[a]	87.45
number trinity transcripts	121 017
number trinity "genes"	72 657
% GC	42.86
N50 value	1579
median contig length	544
average contig length	928
total assembled bases	112 347 148
min transcript size (bp)	201
max transcript size (bp)	11 690
number of transcripts with FPKM ≥ 1.0	61 942

[a]Average of the quadruplicates.

CoA, IPP, and DMAPP (Figures S6–S9, Supporting Information).

With regard to terpenoid building blocks, the transcriptome analysis revealed an interesting fact that complemented the results obtained from the proteomic analysis. While shotgun proteomics detected only the mevalonate pathway in the leaves (Table 1), the transcriptome analysis showed that both the mevalonate and methylerythritol phosphate pathways were active in this organ. It was confirmed by the identification of all the enzymes involved in the production of IPP and DMAPP from both pathways (Figure S9, Supporting Information). This result is in accordance with studies that have suggested the simultaneous operation of these two separate biosynthesis pathways in Piperaceae.[23,24]

As observed in the proteome, the transcriptome analysis identified a large number of proteins involved in terpene biosynthesis (Figure S9–S12, Supporting Information). Important enzymes such as geranyl diphosphate, farnesyl diphosphate, geranylgeranyl diphosphate, and squalene synthases were identified, along with a number of prenyltransferases. Some of these enzymes may be involved in the biosynthesis of the terpenoidal portions present in the chromans isolated from this species. In addition, some enzymes related to the biosynthesis of monoterpenes, sesquiterpenes, diterpenes, triterpenes, and steroids were identified (Table S13, Supporting Information). In the case of diterpenes, the complete biosynthesis pathway for gibberellins was detected (Figure S10, Supporting Information). A similar trend was observed for jasmonic acid, where virtually all enzymes involved

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

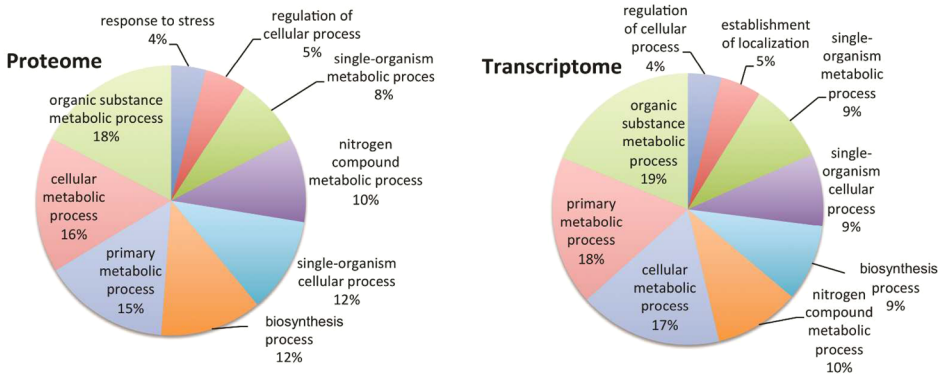


Figure 4. Comparison of gene ontology biological process annotations of the identified proteins and transcripts from the leaves of *P. obtusifolia*.

in the formation of (−)-methyl jasmonate were identified (Figure S11, Supporting Information). These data complement the results obtained with the proteomic analysis.

The transcriptome analysis also allowed the identification of many enzymes involved in the production of phenylpropanoids. Almost all enzymes involved in the early steps, starting from D-erythose-4-phosphate to phenylalanine, until the later steps that lead to the formation of flavonoid and lignoid precursors, were identified (Figure S12, Supporting Information). As for the flavonoids, some enzymes related to the biosynthesis of 4′-*O*-methylisoswertisin-2″-*α*-L-rhamnoside and isoswertisin-2″-*α*-L-rhamnoside, derived from apigenin, were found (Figure S13, Supporting Information). Moreover, enzymes related to isoflavonoid and anthocyanin biosyntheses were also identified. This corroborates the data from the shotgun proteomic analyses.

The combined proteomic and transcriptomic analyses identified various cytochrome P450 enzymes related to the biosynthesis of secondary metabolites in the leaves of *P. obtusifolia*.

While shotgun proteomics allowed the identification of polyketide synthase type III in the leaves, the transcriptome analysis revealed the presence of polyketide cyclase/dehydrase in the same organ. However, the presence of the polyketide synthase related to the biosynthesis of orsellinic acid was not confirmed in either of these two approaches. As orsellinic acid is commonly isolated from fungi and bacteria, a reanalysis of the plant transcripts was carried out by comparing them with the nonredundant protein database of NCBI and with a custom orsellinic acid synthase database (from plants, fungi, and bacteria), including eight domains of these enzymes. Even after using this more in-depth and specific approach, this enzyme remained undetected, which raised the question of whether the plant is actually responsible for the biosynthesis of this compound. Considering that the host plant can metabolize products from the endophyte and *vice versa*,[10] an attempt to identify this enzyme may be based on studies of PKSs from the endophytes intimately associated with *P. obtusifolia*.

Despite not having identified the proteins related to the biosynthesis of orsellinic acid, the transcriptome analysis revealed an interesting fact, i.e., the presence of a number of prenyltransferases and the tocopherol cyclase enzyme (Table S13, Supporting Information). The former enzymes may be related to the prenylation of orsellinic acid, while the latter may be responsible for the cyclization that yields the 3,4-dihydro-2*H*-pyran moiety. It is noteworthy, however, that tocopherol cyclase is supposed to produce benzopyrans stereoselectively[25] in contrast to the racemic compounds found in *P. obtusifolia*. Figure 3 presents all enzymes involved in the biosynthesis of chromans, identified using both proteomic and transcriptomic analyses.

The results of this work showed that the proteomic and transcriptomic techniques were complementary in the identification of the active secondary metabolism pathways in *P. obtusifolia*. In addition, a comparative analysis of the gene ontology biological processes of the proteins and transcripts obtained for the leaves showed similar main subcategories (Figure 4). For this analysis the proteins of both soluble and microsomal fractions of the leaves were considered. Some differences have been observed that may be considered normal since the proteome and transcriptome are dynamic entities and may vary qualitatively and also quantitatively under different conditions.[26]

Phytochemical Profiling. Finally, in order to corroborate the proteomic and transcriptomic results described above, UPLC-DAD-MS analyses of ethanol extracts of *P. obtusifolia* (Figure 5) were performed to confirm the presence of the prenylated benzopyrans (**1**−**4**) as well as the flavonoid glucoside 4′-*O*-methylisoswertisin-2″-*α*-L-rhamnoside (**5**). Furthermore, the presence of the possible acyclic precursors (**6** and **7**) to the chromans was evident.

In conclusion, the present work demonstrated that despite being a CAM plant, *P. obtusifolia* possesses a rich secondary metabolism biosynthesis, as evidenced by the presence of a large number of enzymes involved in the formation of various classes of compounds in its organs. Terpenoid biosynthesis was the pathway with the largest number of enzymes identified. In this matter, the results of the present work showed that both mevalonate and methylerythritol phosphate terpene-production routes are active in the leaves of *P. obtusifolia*. These pathways are related to the biosynthesis and structural diversification (esterification with monoterpenes) of the prenylated chromans (orsellinic acid derivatives), the main secondary metabolites present in this plant. These compounds are extremely interesting from both chemical and biological points of view because of their peculiar biosynthesis and potent trypanocidal activity. Additionally, the combination of the shotgun proteomics and transcriptome analysis allowed for the identification of other active pathways, such as the biosynthesis of flavonoids and phenylpropanoids.

However, the above-mentioned approaches did not identify the polyketide synthases possibly involved in the production of orsellinic acid. This latter fact reinforces the question of whether the plant is actually responsible for the biosynthesis of

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

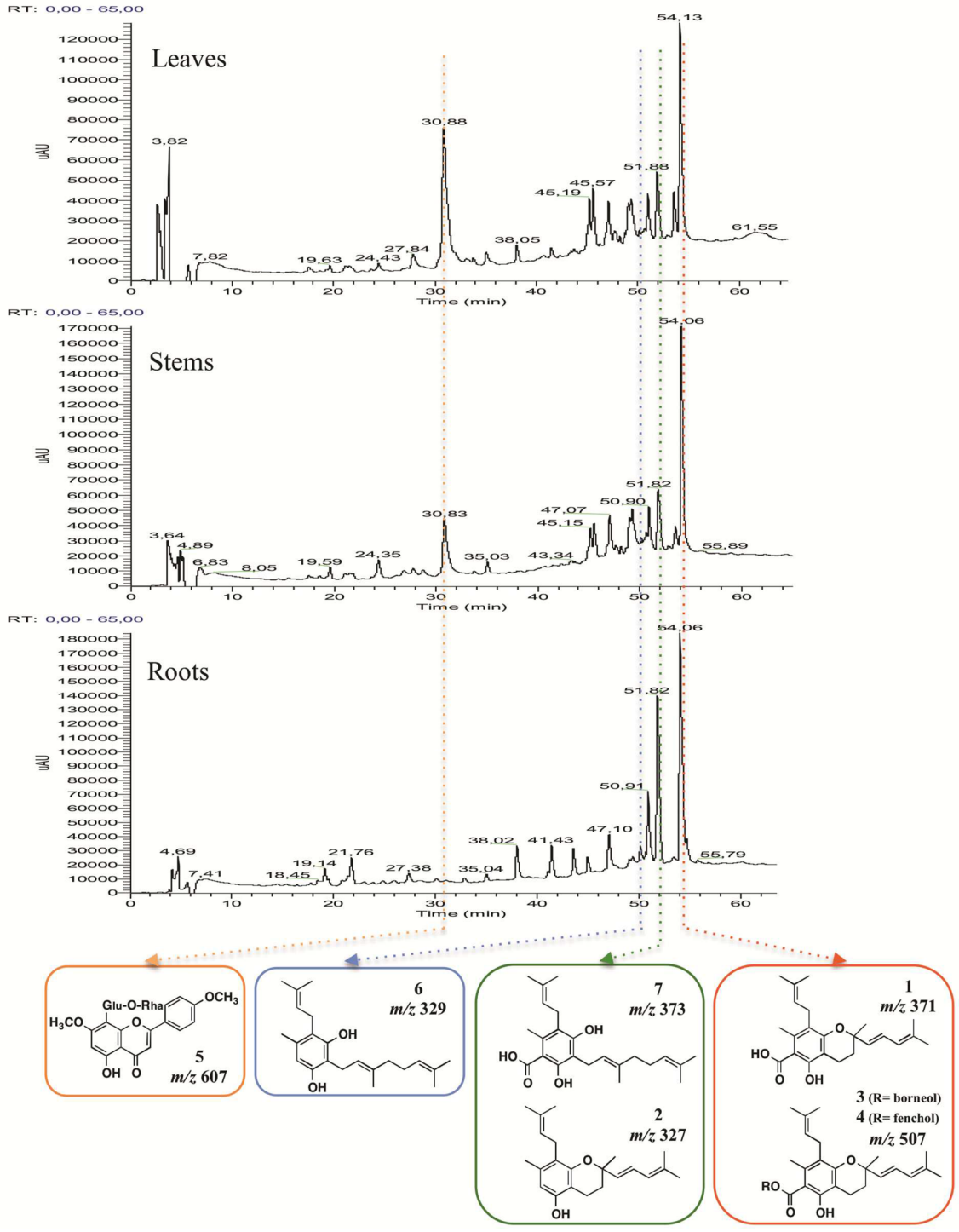


Figure 5. Chromatograms of the EtOH extracts of the leaves, stems, and roots of *P. obtusifolia* obtained using UPLC-DAD-MS.

this class of compounds. Orsellinic acid moieties found in *P. obtusifolia* are, therefore, likely to be produced by endophytes intimately associated with the plant. The transcriptome analysis, however, revealed the expression of the tocopherol cyclase enzyme in the leaves. This enzyme may be responsible for the cyclization of the prenylated orsellinic acid precursor that yields the 3,4-dihydro-2*H*-pyran moiety. These results collectively suggest that orsellinic acid-derived benzopyrans may be formed by the combination of biosynthesis efforts from the host plant and endophytic fungi.

Finally, the results of the present study demonstrate that it is possible to use shotgun proteomic and transcriptomic approaches to investigate the genesis and biochemical transformations of a series of natural products in the nonmodel plant *P. obtusifolia.*

EXPERIMENTAL SECTION

General Experimental Procedures. The LCMS-IT-TOF analyses (MS and MS^n) were performed on a hybrid system, where a UFLC (Prominence) was connected to an IT/TOF MS (ion trap/time-of-flight mass spectrometer) instrument (Shimadzu, Kyoto,

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

Japan), equipped with an electrospray ionization source. The UFLC system consisted of two LC20AD pumps, a SIL-20 AHTautomatic injector, an SPD M20A diode array detector, and a CTO-20A column oven. The LC separation of the EtOH extracts was performed on an ACCELA UPLC system, equipped with a photodiode array detector and an LCQ Fleet ion trap mass spectrometer (Thermo Scientific, Austin, TX, USA). The RNA libraries were sequenced using an Illumina HiSeq2500 device (Illumina, San Diego, CA, USA).

Plant Material. *Peperomia obtusifolia* (L.) A. Dietr. was collected from Araraquara, SP, Brazil, and was identified by Dr. Inês Cordeiro (Instituto de Botânica, São Paulo, SP, Brazil). The voucher specimen (Kato-070) was deposited at the Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" (São Paulo, SP, Brazil). Specimens were cultivated under greenhouse conditions at the Instituto de Química-UNESP, Araraquara.

Protein Extraction. Leaves, stems, and roots of *P. obtusifolia* were harvested from an adult specimen and immediately frozen in liquid N_2. The frozen material (12 g) was ground in an analytical mill with the addition of 60 mL of 10 mM NH_4OAc (pH 6.8) containing 20% polyvinylpolypyrrolidone (PVPP), 10 mM dithiothreitol (DTT), 1 mM ethylenediaminetetraacetic acid (EDTA), and Sigma*FAST* (Sigma, Saint Louis, MO, USA) protease inhibitor cocktail tablet (1 tablet can be used to prepare 100 mL of solution). The homogenate was stirred slowly for 15 min in an ice bath, filtered on gauze, and centrifuged at 10 000 rpm for 10 min.

Fifty microliters of the supernatant was subjected to ultracentrifugation at 39 000 rpm for 1 h. The resulting pellets were suspended in 30 mL of 100 mM Na_2CO_3 (pH 11.5). The suspensions were subjected to ultracentrifugation as described above. The supernatants (soluble fractions) were subjected to protein determination and the shotgun proteomic analyses, and the pellets were suspended in a buffer containing 3 mL of 10 mM DTT, 1 mM EDTA, and Sigma*FAST* protease inhibitor cocktail tablet (1 tablet per 100 mL of solution) in 10 mM NH_4OAc (pH 6.8) and sonicated (4.4 s "on" 40% amplitude/9 s "off", repeated 27 times). The resulting solutions (microsomal fractions) were also submitted to protein determination and the shotgun proteomic analyses.

Protein Determination. The Bradford modified method was used for the determination of protein concentrations, with bovine serum albumin serving as the standard.[27]

In-Solution Digestion. In-solution digestion was performed to prepare the samples for the shotgun analysis. For this purpose, the protein extracts (500 μg for soluble fractions and 100 μg for microsomal fractions) were solubilized using 50 mM NH_4HCO_3 (pH 7.9) containing 7.5 M urea, maintained at room temperature for 1 h, and reduced using 10 mM DTT at 37 °C for 1 h. Afterward, the proteins were alkylated using 40 mM iodoacetamide at 25 °C for 1 h in the dark. The samples were diluted 5-fold with 100 mM NH_4HCO_3, pH 7.8, and 1 M $CaCl_2$ was added to the samples to a final concentration of 1 mM. Nonautolytic trypsin (Promega, Madison, WI, USA) was added to the denatured protein solution (1:50 trypsin:protein, w/w), and the reaction mixture was incubated for 18 h at 37 °C. The digested samples were desalted using a Sep-Pak C_{18} 1 cm^3 Vac Cartridge column (Waters, Milford, MA, USA). The columns were conditioned with 0.1% trifluoroacetic acid (TFA) in MeCN. The peptides were eluted from the column with 1 mL of 0.1% TFA/80% MeCN and concentrated in a Speed-Vac system to dryness. The digested samples were stored at 80 °C until analysis; the tryptic peptides were solubilized in 50% MeCN and subjected to LCMS-IT-TOF/MS^n analysis.

LCMS-IT-TOF Analyses. The LC system was connected to a Shim-pack XR-ODS C_{18} column (3 × 100 mm, 120 Å, 2.2 μm) (Shimadzu, Kyoto, Japan), and a gradient of MeCN from 5% (v/v) to 95% (v/v) over 180 min with a flow rate of 0.2 mL/min was used for elution. The elution was monitored at 214 nm, and the eluates were analyzed in the positive ion mode. The column was directly connected to the ionization system of the mass spectrometer, which was set to permit the accumulation of all ions in the octupole, followed by rapid pulsing into the IT for MS^n analysis and then the introduction into the TOF sector for accurate mass determinations. The conditions set for optimal operation were as follows: positive mode, an electrospray voltage of 4.5 kV, a CDL temperature of 200 °C, a block heater temperature of 200 °C, a nebulizer gas (N_2) flow of 1.5 L/min, a trap cooling gas (Ar) flow of 95 mL/min, an ion trap pressure of 1.7×10^{-2} Pa, a TOF region pressure of 1.5×10^{-4} Pa, an ion accumulation time of 50 ms, a collision energy set at 35% for both MS2 and MS3, and a collision gas set at 20%. Autotuning of the instrument was performed with a Na-TFA solution using the following parameters: for the positive mode, an error of 3.08 ppm and a resolution of 15.414; for the negative mode, an error of 2.69 ppm and a resolution of 13.212.

Protein Identification. MASCOT searches were conducted using MASCOT 2.2 (Matrix Science, London, UK) against the publicly available nonredundant protein sequence Swiss-Prot database (on 09/17/2014). It must be emphasized that there is no genomic information available in the protein data banks for *P. obtusifolia*; however, plant proteins still may be identified by using cross-species data.[13] For this purpose, all 4 098 531 entries contained in the taxa Viridiplantae were selected. Search parameters were set as follows: taxonomy, trypsin as the enzyme, two maximum missing cleavage sites allowed, a peptide mass tolerance of 0.2 Da for the MS and 0.2 Da for the MS/MS spectra, carbamidomethyl (C) as a fixed modification, and methionine oxidation specified as variable modifications. The FDR (false discovery rate) was calculated using the original decoy FDR approach from Mascot; a separate decoy database was generated from the protein sequence database with the decoy.pl Perl script provided by Matrix Science. This script randomizes each entry, while it retains the average amino acid composition and length of the entries. For protein identification, the protein and peptide FDR rates were selected between 0.1 and 0.01. FDR was calculated based on the Mascot Score. After these procedures, there was a refinement of peptide identification by using the Protein Prophet algorithm; the identifications were accepted if they could be established with ≥99.0% probability. To understand and interpret the data from the shotgun proteomic approach, the proteins were assigned to gene ontology (GO) terms according to UniProt Knowledgebase (UniProtKB)[28] and mapped against the GO database by using the Blast2GO program.[29] In this latter analysis, the proteins may have been annotated to more than one category, since the same protein may be involved in several biological processes.

RNA Isolation, Library Preparation, and Sequencing. Total RNA was extracted from the leaves of *P. obtusifolia* using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and DNase I was used to remove contaminating DNA. The quality of isolated RNA was checked in a 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The extraction was performed in quadruplicate following the manufacturer's instructions. The paired-end library of the four samples was prepared using an Illumina TruSeq RNA sample preparation kit, and the libraries were sequenced using an Illumina HiSeq2500 device.

***De Novo* Transcriptome Assembly, Annotation, and Analysis.** About 3 GB of sequence data was obtained for each library, which correspond to an average of 36 330 714 raw reads. To consistently ensure quality and adapter trimming to the sequences, quality control was performed using the Trimmomatic software, version 0.33.[30] The parameters considered for filtration were adapter trimming, cutting the average quality read (threshold of 15), and removal of bases of the start of a read and the end of a read (score < 3) along with dropping off the reads below 50 bp length. After the quality assessment, the clean sequences were an average of 31 888 176, which reflected a loss of ∼13%. Digital normalization was performed using the Diginorm software, version 1.4+28.g6f18434,[31] thereby decreasing sampling variation, discarding redundant data, and removing the majority of errors. The reduced data set was assembled using the Trinity software, version r20140717,[32] on default parameters with mismatch cost 2 and insertion and deletion cost 3. The assembled sequences were deposited at http://nubbe.iq.unesp.br/portal/do/Query?service=23&name=transcripts.peperomia.fasta.zip.

Fragments per kilobase of transcript per million mapped reads (FPKM) values were calculated by first aligning the trimmed reads to the assembled transcriptome using the Bowtie2 program.

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

To annotate the assembled transcripts, the *ab initio* assembly was compared with (1) a custom protein database with a total of 948 000 sequences from plant proteins, derived from public banks RefSeq and UniProt/Swiss-Prot; (2) a nonredundant protein database of NCBI, and (3) a custom protein database with a total of 45 sequences from orsellinic acid synthase proteins from NCBI (from plant, fungi, and bacteria) and eight domains of these enzymes obtained from Pfam (http://pfam.xfam.org),[33] using BLASTx with an E-value threshold of 1×10^{-5}. The GO classifications of the annotated transcripts against the custom plant protein database were conducted by using the Blast2GO program.[29] Using the KEGG BlastKOALA annotation Web tool, these annotated transcripts were searched against a nonredundant set of KEGG GENES from Kyoto Encyclopedia of Genes and Genomes (KEGG) database to assign the K numbers.[34] Finally, experimental data from the transcriptomic analysis were mapped using the search and color pathway tool offered by the KEGG database (http://www.genome.jp/kegg/tool/map_pathway2.html).[35]

Extraction of Secondary Metabolites and UPLC-DAD-MS Analyses. For the qualitative analysis of the extracts of secondary metabolites, powdered leaves, stems, and roots from an adult plant were extracted with EtOH in the dark at room temperature for 3 h. The solvent was evaporated under reduced pressure without heating, and the samples were promptly injected into the UPLC system. A Phenomenex SYNERGI Hydro-PR-80A (250 × 4.6 mm, 4 μm) column (Torrance, CA, USA) was used for the separation, and all solvents were of HPLC or analytical grade. The chromatographic profiling employed a mobile phase consisting of 0.1% aqueous HOAc (solvent A) and MeOH (solvent B) with linear gradient elution from 5% to 100% B, in 50 min, followed by 15 min of 100% of B, at a flow rate of 1.0 mL/min. The eluate was monitored at 254 nm, and the mass analyses were performed in the positive ion mode.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.6b00827.

(Protein identification data (accession number, peptide sequence, ion score, protein score, % of coverage, functional category distribution), transcriptome data (EC number, KO, FPKM, secondary metabolism pathways), secondary metabolism proteins (schemes and KEGG maps) PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*Tel (A. N. L. Batista): +55-16-3301-9508. Fax: +55-16-3322-2308. E-mail: andrluca@yahoo.com.br.

*Tel (M. Furlan): +55-16-3301-9661. Fax: +55-16-3322-2308. E-mail: maysaf@iq.unesp.br.

ORCID

Andrea N. L. Batista: 0000-0003-4185-9316

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank São Paulo Research Foundation (FAPESP grants 2011/51684-1, CIBFar 2013/07600-3, and 2014/25222-9) for funding. A.N.L.B. thanks FAPESP for the provision of a fellowship (2011/01003-8). We are also grateful to the National Council for Scientific and Technological Development (CNPq) for research fellowships.

■ REFERENCES

(1) Tanaka, T.; Asai, F.; Iinuma, M. *Phytochemistry* **1998**, *49*, 229−232.

(2) Lentz, D. L.; Clark, A. M.; Hufford, C. D.; Meurer-Grimes, B.; Passreiter, C. M.; Cordero, J.; Ibrahimi, O.; Okunade, A. L. *J. Ethnopharmacol.* **1998**, *63*, 253−263.

(3) Mota, J. S.; Leite, A. C.; Batista, J. M., Jr; López, S. N.; Ambrósio, D. L.; Passerini, G. D.; Kato, M. J.; Bolzani, V. S.; Cicarelli, R. M. B.; Furlan, M. *Planta Med.* **2009**, *75*, 620−623.

(4) Batista, J. M., Jr; Batista, A. N. L.; Mota, J. S.; Cass, Q. B.; Kato, M. J.; Bolzani, V. S.; Freedman, T. B.; López, S. N.; Furlan, M.; Nafie, L. A. *J. Org. Chem.* **2011**, *76*, 2603−2612.

(5) López, S. N.; Lopes, A. A.; Batista, J. M., Jr.; Flausino, O., Jr.; Bolzani, V. S.; Kato, M. J.; Furlan, M. *Bioresour. Technol.* **2010**, *101*, 4251−4260.

(6) Batista, J. M., Jr; Batista, A. N. L.; Rinaldo, D.; Vilegas, W.; Cass, Q. B.; Bolzani, V. S.; Kato, M. J.; López, S. N.; Furlan, M.; Nafie, L. A. *Tetrahedron: Asymmetry* **2010**, *21*, 2402−2407.

(7) Batista, J. M., Jr; Batista, A. N. L.; Kato, M. J.; Bolzani, V. S.; López, S. N.; Nafie, L. A.; Furlan, M. *Tetrahedron Lett.* **2012**, *53*, 6051−6054.

(8) Cox, R. *Org. Biomol. Chem.* **2007**, *5*, 2010−2026.

(9) Alvin, A.; Miller, K. I.; Neilan, B. A. *Microbiol. Res.* **2014**, *169*, 483−495.

(10) Ludwig-Müller, J. *Biotechnol. Lett.* **2015**, *37*, 1325−1334.

(11) Batista, J. M., Jr.; Lopez, S. N.; Mota, J. S.; Vanzolini, K. L.; Cass, Q. B.; Rinaldo, D.; Vilegas, W.; Bolzani, V. S.; Kato, M. J.; Furlan, M. *Chirality* **2009**, *21*, 799−801.

(12) Swanson, S. K.; Washburn, M. P. *Drug Discovery Today* **2005**, *10*, 719−725.

(13) Liska, A. J.; Shevchenko, A. *Proteomics* **2003**, *3*, 19−28.

(14) Góngora-Castillo, E.; Buell, C. R. *Nat. Prod. Rep.* **2013**, *30*, 490−500.

(15) Mirzaei, M.; Soltani, N.; Sarhadi, E.; Pascovici, D.; Keighley, T.; Salekdeh, G. H.; Haynes, P. A.; Atwell, B. J. *J. Proteome Res.* **2012**, *11*, 348−358.

(16) Mühlenweg, A.; Melzer, M.; Li, S.; Heide, L. *Planta* **1998**, *205*, 407−413.

(17) Sattler, S. E.; Cahoon, E. B.; Coughlan, S. J.; DellaPenna, D. *Plant Physiol.* **2003**, *132*, 2184−2195.

(18) Yang, D. L.; Yao, J.; Mei, C. S.; Tong, X. H.; Zeng, L. J.; Li, Q.; Xiao, L. T.; Sun, T. P.; Li, J.; Deng, X. W.; Lee, C. M.; Thomashow, M. F.; Yang, Y.; He, Z.; He, S. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1192−200.

(19) Mota, J. S.; Leite, A. C.; Kato, M. J.; Young, M. C.; Bolzani, V. S.; Furlan, M. *Nat. Prod. Res.* **2011**, *25*, 1−7.

(20) Hu, L.; Hao, C.; Fan, R.; Wu, B.; Tan, L.; Wu, H. *PLoS One* **2015**, *10*, e0129822.

(21) Shi, Y.; Yan, X.; Zhao, P.; Yin, H.; Zhao, X.; Xiao, H.; Li, X.; Chen, G.; Ma, X. F. *PLoS One* **2013**, *8*, e63993.

(22) Long, Y.; Zhang, J.; Tian, X.; Wu, S.; Zhang, Q.; Zhang, J.; Dang, Z.; Pei, X. W. *BMC Genomics* **2014**, *15*, 1111.

(23) Lopes, A. A.; Baldoqui, D. C.; López, S. N.; Kato, M. J.; Bolzani, V. S.; Furlan, M. *Phytochemistry* **2007**, *68*, 2053−2058.

(24) Leite, A. C.; Lopes, A. A.; Kato, M. J.; Bolzani, V. S.; Furlan, M. *J. Braz. Chem. Soc.* **2007**, *18*, 1500−1503.

(25) Grütter, C.; Alonso, E.; Chougnet, A.; Woggon, W. *Angew. Chem., Int. Ed.* **2006**, *45*, 1126−1130.

(26) Twyman, R. M. *Principles of Proteomics*; Garland Science: New York, 2013.

(27) Stoscheck, C. M. In *Methods in Enzymology*; Deutscher, M. P., Ed.; Academic Press: San Diego, 1990; Vol. *182*, Chapter 6, pp 50−68.

(28) Huntley, R. P.; Sawford, T.; Mutowo-Meullenet, P.; Shypitsyna, A.; Bonilla, C.; Martin, M. J.; O'Donovan, C. *Nucleic Acids Res.* **2015**, *43* (Database issue), D1057−1063.

(29) Conesa, A.; Götz, S. *Int. J. Plant Genomics* **2008**, *2008*, 1−13.

(30) Bolger, A. M.; Lohse, M.; Usadel, B. *Bioinformatics* **2014**, *30*, 2114−2120.

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX

(31) Brown, C. T.; Howe, A.; Zhang, Q.; Pyrkosz, A. B.; Brom, T. H. 2012, *arXiv:1203.4802* [q-bio.GN], http://arxiv.org/abs/1203.4802.
(32) Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. *Nat. Protoc.* **2013**, *8*, 1494−1512.
(33) Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; Salazar, G. A.; Tate, J.; Bateman, A. *Nucleic Acids Res.* **2016**, *44*, D279−D285.
(34) Kanehisa, M.; Sato, Y.; Morishima, K. *J. Mol. Biol.* **2016**, *428*, 726−731.
(35) Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. *Nucleic Acids Res.* **2016**, *44*, D457−462.

DOI: 10.1021/acs.jnatprod.6b00827
J. Nat. Prod. XXXX, XXX, XXX−XXX