



Biosynthetic Insights into *p*-Hydroxybenzoic Acid-Derived Benzopyrans in *Piper gaudichaudianum*

Andrea N. L. Batista,^{*a} João M. Batista Jr.,^b Tatiana M. Souza-Moreira,^c Sandro R. Valentini,^c
Massuo J. Kato,^d Cleslei F. Zanelli^c and Maysa Furlan^{*a}

^aInstituto de Química, Universidade Estadual Paulista (UNESP), 14800-060 Araraquara-SP, Brazil

^bDepartamento de Química, Universidade Federal de São Carlos (UFSCar), 13565-905 São Carlos-SP, Brazil

^cFaculdade de Ciências Farmacêuticas, Universidade Estadual Paulista (UNESP),
14801-902 Araraquara-SP, Brazil

^dInstituto de Química, Universidade de São Paulo (USP), 05508-000 São Paulo-SP, Brazil

Piper gaudichaudianum Kunth (Piperaceae) accumulates gaudichaudianic acid, a prenylated benzopyran, as its major component. Interestingly, this trypanocidal compound occurs as a racemic mixture. Herein, transcriptomic investigations of *Piper gaudichaudianum* using the RNA-seq approach are reported, and from the analysis of the transcripts expressed it was possible to propose a complete biosynthetic pathway for the production of gaudichaudianic acid, including the steps that originate its precursor, *p*-hydroxybenzoic acid. *Peperomia obtusifolia* (L.) A. Dietr. (Piperaceae) also accumulates racemic benzopyrans, however, its chromanes originate from the polyketide pathway, while the chromenes from *Piper* derives from the shikimate pathway. Recent transcriptomic and proteomic studies of the former species did not identify polyketide synthases involved in the production of the benzopyran moiety, but revealed the expression of tocopherol cyclase, which may be responsible for the cyclization of the 3,4-dihydro-2*H*-pyran ring. The analysis of the enzymes involved in the secondary metabolism of *Piper gaudichaudianum* and the comparison with the data previously obtained from *Peperomia obtusifolia* can provide valuable information on how these compounds are biosynthesized.

Keywords: Piperaceae, chromene, transcriptome, RNA-seq, *de novo* assembly, biosynthesis

Introduction

Chromanes and chromenes, which are secondary metabolites commonly found in some species from *Piper* and *Peperomia* genera (Piperaceae), are characterized by the presence of a benzopyran ring with various levels of oxidation.¹⁻³ This nucleus is considered a privileged structure that is very common in bioactive natural products, such as coumarins, flavones, tocopherols (vitamin E) and tetrahydrocannabinoids.^{4,5}

The benzopyran moieties, 2*H*-chromenes and chromanes, isolated from *Piper gaudichaudianum* and *Peperomia obtusifolia*, respectively, have been demonstrated as potent trypanocidal compounds.^{6,7} Curiously, both classes of compounds occur as racemic mixtures in these species, though their formation follows two distinct biosynthetic routes.^{6,8,9} In *Piper gaudichaudianum*,

chromenes originate from the shikimate pathway and use *p*-hydroxybenzoic acid (*p*-HBA) as a precursor.¹⁰ In the case of *Peperomia obtusifolia*, chromanes are formed through the polyketide pathway and use orsellinic acid as a precursor.^{1,11} Additionally, the formation of both classes of metabolites involves the condensation of an aromatic unit (*p*-HBA or orsellinic acid) and an isoprene unit (dimethylallyl pyrophosphate, isopentenyl pyrophosphate or geranyl pyrophosphate), followed by cyclization that gives rise to the benzopyran ring. During cyclization, the carbon atom at C-2 becomes a stereogenic center and, different from other benzopyrans such as vitamin E, both enantiomers are biosynthesized (Figure 1). Thus, the study of the proteins and genes possibly responsible for the biosynthesis of benzopyrans in these species, as well as the comparison between them, may provide new insights into how these compounds are produced.

Recent studies in *Peperomia obtusifolia* using a combination of shotgun proteomics and transcriptome analysis did not identify an orsellinic acid synthase.

*e-mail: andrluca@yahoo.com.br; maysaf@iq.unesp.br

This paper is part of the PubliSBQ Special Issue "IUPAC-2017" (<http://publi.s bq.org.br/>).



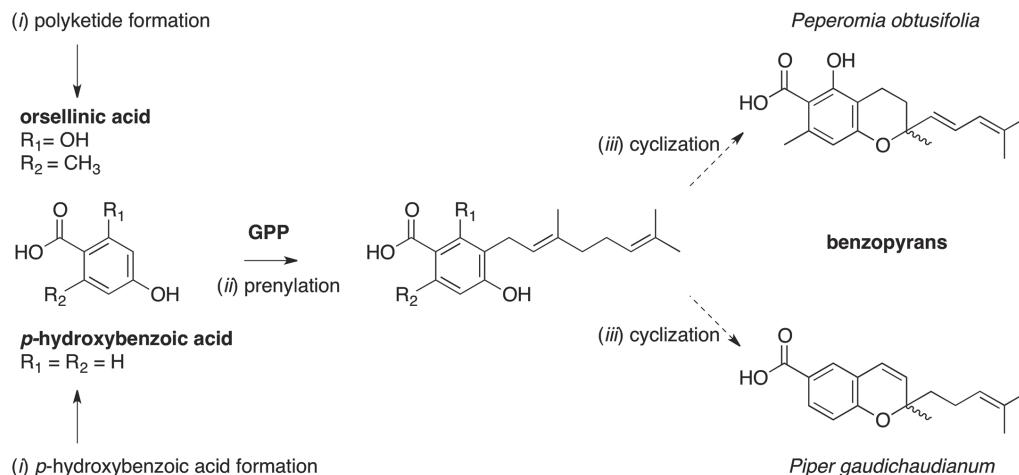


Figure 1. The proposed biosynthetic pathway of benzopyrans from *Peperomia obtusifolia* and *Piper gaudichaudianum*.

However, the transcriptome analysis revealed the expression of tocopherol cyclase that may be responsible for the cyclization of the prenylated orsellinic acid precursor that yields the 3,4-dihydro-2*H*-pyran ring. Because orsellinic acid is commonly found in sordariomycetes and eurotiomycetes, these results suggest that orsellinic acid-derived benzopyrans may be formed by the combination of biosynthetic efforts from the host plant and endophytic fungi.¹²

Some studies have already been performed in *Piper gaudichaudianum* aimed at elucidating the biosynthesis of gaudichaudianic acid (**1**) (Figure 2),¹⁰ which is the major prenylated chromene isolated from this species and described as a potent trypanocidal compound against the Y-strain of *Trypanosoma cruzi*.¹³ Interestingly, trypanocidal assays indicated that the (+)-enantiomer was more active than its antipode and that the enantiomer mixtures showed a synergistic effect, with the racemic mixture being the most active.⁶ Regarding its biosynthesis, the origin of the terpene moieties was shown to involve both the mevalonate and methylerythritol phosphate pathways.¹⁰

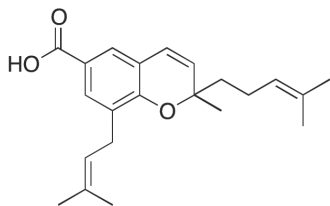


Figure 2. Chemical structure of the gaudichaudianic acid.

Thus, to complement the data in the literature and elucidate the complete biosynthesis of chromenes from *Piper gaudichaudianum*, a transcriptome study was performed on leaves from this plant using the RNA-seq approach.

Results and Discussion

Transcriptome sequencing, *de novo* assembly and annotation

The racemic prenylated chromene gaudichaudianic acid (**1**) is the major constituent in all organs in *Piper gaudichaudianum* adult plants, though it is found exclusively in the roots of seedlings.⁶ The occurrence of this compound has been reported in leaves over 12 months of age, and it becomes the major compound by the 15th month of growth.¹⁴ Thus, to obtain the *Piper gaudichaudianum* transcriptome, three RNA-seq libraries were constructed from a pool of leaves from young plants (older than 15 months) and sequenced using the HiSeq 2500 Illumina paired-end sequencing system. RNA-seq is an approach to transcriptome profiling that uses deep-sequencing technologies in which RNA is sequenced via a high-throughput manner enabling robust assessments of eukaryotic transcriptomes.^{15,16} Furthermore, paired-end sequencing produces twice the number of reads at the same time, and sequences aligned as read pairs enable more accurate read alignment.¹⁷ This approach has been particularly useful in non-model species.¹⁶

In this study, an average of 34,846,653 raw reads with a length of 100 bp each and a percentage of bases with Q (Phred quality score) ≥ 30 over 87.5% were generated from the three samples submitted to Illumina sequencing. These values are adequate to guarantee high-quality data. When the sequencing quality reaches Q30, virtually all the reads will be perfect and have zero errors and ambiguities. After removing the adaptors and low-quality reads, 30,449,535 clean reads were retained, which reflected a loss of ca. 12.6%. These clean data were *de novo* assembled into a total of 216,786 transcripts with

average length of 770 bases and an N50 length (weighted metric that represents the minimum assembly size in which $\geq 50\%$ of the assembled bases are found)¹⁶ of 1,362 bases using the Trinity software,¹⁸ which generates transcript contigs based on overlapping information. The transcript size distribution revealed that there were 165,192 contigs (76.2%) ranging from 100 to 1,000 bases, 32,880 contigs (15.2%) ranging from 1,001 to 2,000 bases, and 18,701 contigs (8.6%) longer than 2,000 bases (Figure S1, Supplementary Information). The details of the transcript assembly are summarized in Table 1. The present results are in accordance with those recently published for transcriptomes for the Piperaceae species *Piper nigrum* and *Peperomia obtusifolia*, which were assembled using Illumina HiSeq 2000 and 2500 platforms, respectively.^{12,19}

Table 1. Overview of the sequencing and *de novo* assembly of *Piper gaudichaudianum* transcriptome

Description	<i>Piper gaudichaudianum</i> (leaves)
Total raw reads ^a	34,846,653
Total clean reads ^a	30,449,535
Bases \geq Q30 ^a / %	87.5
Number of trinity transcripts	216,786
Number of trinity 'genes'	148,419
GC / %	46.04
N50 value	1,362
Median contig length	407
Average contig length	771
Total assembled bases	167,094,492

^aAverage of the triplicate. Q30: Phred quality score of 30; GC: guanine-cytosine content; N50: minimum contig length needed to cover 50% of assembled bases.

Sequence homology searches were conducted for the transcript functional annotation. Therefore, all transcripts were submitted to BLASTX²⁰ searches against the NCBI non-redundant (Nr) protein database, and 148,113 (68%) transcripts had significant similarity with an E (expect) value less than $1e^{-5}$ (Table 2). The sequences were also searched against a custom plant protein database, resulting in 113,374 (52%) annotated transcripts (E values lower than $1e^{-5}$) (Table 2). Even after these two different searches, a significant number of transcripts (68,401, 32%) remained unannotated, which can be a great resource for the discovery of novel genes. This is an expected situation given the absence of genomic information for *Piper gaudichaudianum*.

Table 2. Summary of sequence annotation for *Piper gaudichaudianum*

Database	Hit number (percentage)
NCBI Nr	148,113 (68%)
Custom plant	113,374 (52%)

To classify the functions of the predicted transcripts, functional annotations were also created by using the Blast2GO²¹ software with the Gene Ontology (GO) database, which is an internationally standardized gene functional classification system. A total of 65,579 transcripts were assigned at least one GO term distributed among the three main ontologies comprising biological process, molecular function, and cellular component. The biological process ontology distribution (level 2 of detail) contained mainly proteins involved in metabolic and cellular processes (Figure 3). In the same level of detail, catalytic activity and binding in the molecular function

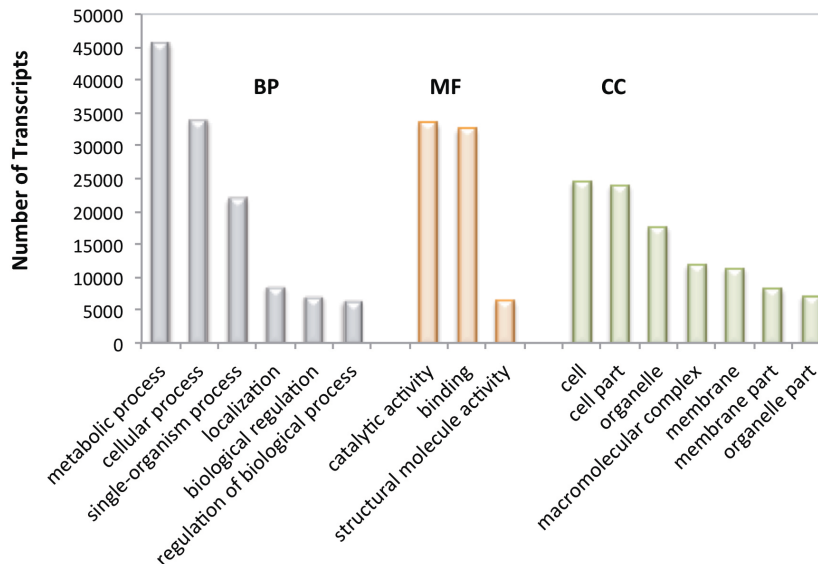


Figure 3. Distribution of GO classifications (level 2). Annotated transcripts were classified into 3 major categories (biological processes (BP), cellular components (CC), and molecular function (MF)).

represented the major subcategories (Figure 3). For cellular components, the assignments were mostly given to cell and cell parts (Figure 3). This pattern of GO annotation distribution was similar to those of other species and is typically seen in the transcriptome of samples undergoing development processes.^{19,22}

KEGG pathway mapping assignment

The annotated transcripts were also analyzed by searching them against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database for KEGG (K) number assignments and subsequent reconstruction of the biosynthetic pathways active in *Piper gaudichaudianum* leaves. KEGG is an integrated database resource for the biological interpretation of genome sequences and other high-throughput data.²³ As a result, 35,192 transcripts were identified with a K number corresponding to a total of 3,594 distinct genes. Among them, 961 (26.7%) are involved in general metabolic pathways that include energy metabolism, carbohydrate and lipid metabolism, nucleotide and amino acid metabolism, and secondary metabolism subcategories.

Within the secondary metabolism subcategory, 423 genes were identified. However, only the transcripts with an FPKM (fragments *per* kilobase of transcript *per* million fragments mapped) ≥ 1 were considered, resulting in 393 genes encoding enzymes involved in some manner in secondary metabolism (Table S1, Supplementary Information). This result allowed for the identification of the active secondary metabolism pathways in *Piper gaudichaudianum*; terpenoid and steroid biosynthesis (61 genes) represented the largest group, followed by phenylpropanoid biosynthesis (18 genes), flavonoid biosynthesis (12 genes), and isoquinoline alkaloid biosynthesis (7 genes). The terpenoid and steroid group includes terpenoid backbone, monoterpene, sesquiterpene, diterpene, triterpene, steroid, and carotenoid biosynthesis. These results are in accordance with the data in the literature because terpenoids, mainly those present in the essential oil, and steroids have already been described in this species.²⁴⁻²⁶ Although there are reports of phenylpropanoids only in the leaves of seedlings,^{14,27} the transcriptome indicates their occurrence in the young plant. These data may suggest that these compounds are produced and then immediately used as precursors of lignins, since the complete biosynthesis pathways of *p*-hydroxyphenyl, guaiacyl, 5-hydroxyguaiacyl, and syringyl lignins have been identified (Figure S2, Supplementary Information). These phenylpropanoids may also act as precursors for the chromenes, flavokawains and flavonoids reported in

this species.^{26,28,29} With respect to isoquinolinic alkaloids, the genes identified are involved in the early steps of their biosynthesis involving dopamine production (Figure S3, Supplementary Information). Although there are no reports of the occurrence of these compounds in *Piper gaudichaudianum*, two aporphinoid alkaloids, cepharadione A and piperolactam E, were isolated from *Piper caninum* and *Piper taiwanense*, respectively.³⁰

Candidate genes involved in benzopyran biosynthesis

In *Piper gaudichaudianum*, chromenes originate from the precursor *p*-HBA via the shikimate pathway. Despite the simple structure of *p*-HBA and its widespread distribution in plants, the enzymatic steps for its biosynthesis are not clearly understood.³¹ Hydroxybenzoates have been reported to originate directly from shikimate via chorismic acid or from phenylalanine (Figure 4).^{31,32}

In the former case, chorismate-pyruvate lyase (CPL) converts chorismic acid into *p*-HBA.³² Although some reports mention that this enzyme is restricted to bacterial lineages,^{33,34} a chorismate pyruvate lyase-catalyzed reaction, similar to that observed in *Escherichia coli*, seems to also occur in eukaryotic microorganisms.^{32,35} However, as expected, CPL was not identified in the *Piper gaudichaudianum* leaf transcriptome, suggesting that *p*-HBA is not biosynthesized by this pathway. This result is in agreement with literature data on microorganisms being the only known organisms to transform chorismic acid into *p*-HBA.

In the latter case, the biosynthesis of *p*-HBA from phenylalanine may follow two different routes, either by means of a β -oxidative or non- β -oxidative pathway.³⁶ The first route proceeds via *p*-coumaroyl-CoA and *p*-hydroxybenzoyl-CoA for the formation of *p*-HBA from *p*-coumaric acid. This pathway involves the activation of *p*-coumaric acid by the action of the enzyme 4-coumarate-CoA ligase (4CL), leading to its thioester with subsequent chain-shortening into *p*-hydroxybenzoyl-CoA in a reaction mechanism analogous to that of NAD-dependent β -oxidation of fatty acids.^{37,38} The operation of this pathway is supported by *in vitro* enzymatic activity assays using cell-free extracts from *Lithospermum erythrorhizon* cell cultures.³⁷ However, this conversion requires steps not fully elucidated yet, i.e., those corresponding to the hydration, dehydrogenation, and thiolation of the β -oxidative cycle followed by the final hydrolysis of the 4-hydroxybenzoyl-CoA thioester.³³ A *Petunia* gene encoding the bifunctional peroxisomal enzyme cinnamoyl-CoA hydratase-dehydrogenase (CHD), which is responsible for the initial two-step conversion of cinnamoyl-CoA into benzoic acid, was

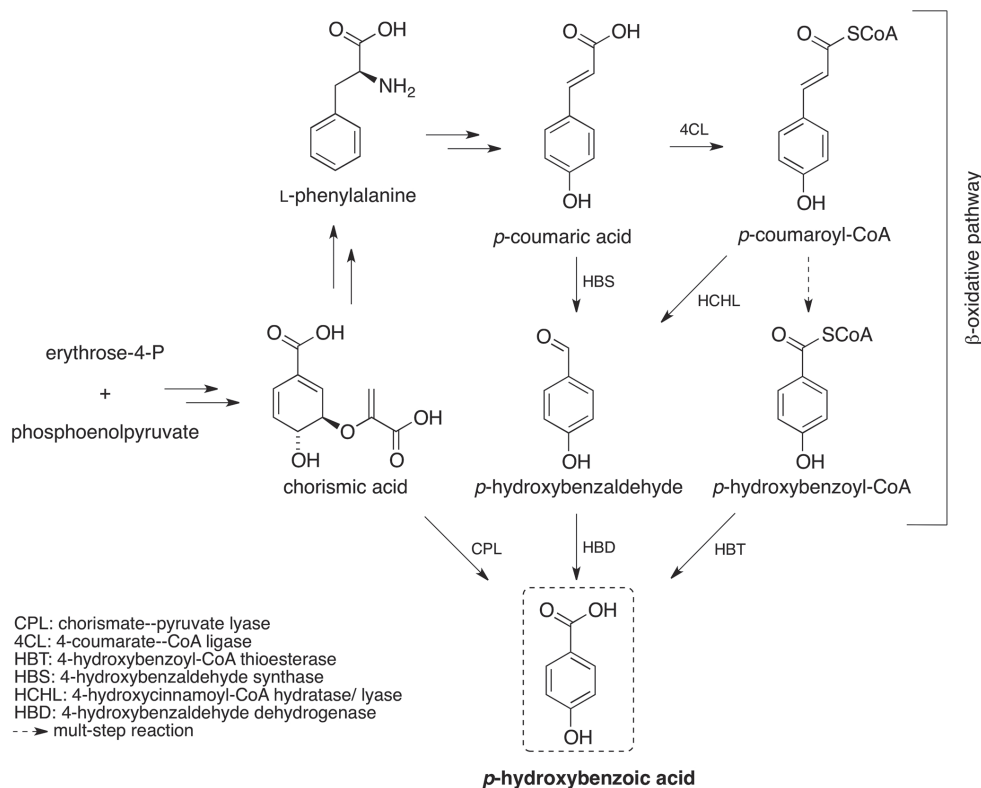


Figure 4. Schematic representation of the possible pathways for the biosynthesis of the *p*-hydroxybenzoic acid.

identified by a functional genomics approach and has been shown to be active with *p*-coumaroyl-CoA (Figure S4, Supplementary Information).³⁹ The next steps are expected to be catalyzed by thiolases and CoA thioesterases. A 3-ketoacyl-CoA thiolase (*PhKAT1*) was confirmed to be involved in the production of benzoyl-CoA from cinnamoyl-CoA in *Petunia hybrid*.³⁶ The specific 4-hydroxybenzoyl-CoA thioesterase (4HBT), which is part of the bacterial 2,4-dichlorobenzoate degradation pathways, appears to occur only in microorganisms. However, CoA thioesterases members of the 4HBT family (1,4-dihydroxy-2-naphthoyl-CoA thioesterase 1 and 2) capable of hydrolyzing aromatic acyl-CoA substrates, including benzoyl-CoA, have been identified in *Arabidopsis*.⁴⁰ The second route involves a non-oxidative pathway for the conversion of *p*-coumaric acid or *p*-coumaroyl-CoA to *p*-hydroxybenzaldehyde in a retro-aldol reaction with no co-factor requirement.^{37,38,41} A 4-hydroxybenzaldehyde synthase (HBS) and 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) catalyzes the penultimate step of *p*-HBA biosynthesis by performing the phenylpropanoid side-chain cleavage of *p*-coumaric acid and *p*-coumaroyl-CoA, respectively.^{31,41} Then, the biosynthesis of *p*-HBA proceeds via *p*-hydroxybenzaldehyde in a reaction catalyzed by the enzyme 4-hydroxybenzaldehyde dehydrogenase.³¹ Studies

with *Daucus carota* and *Vanilla planifolia* support the operation of this pathway.^{31,38}

Analysis of the transcriptome of *Piper gaudichaudianum* leaves against the NCBI non-redundant and custom protein databases allowed for the identification of all the enzymes involved in the biosynthesis of *p*-coumaric acid and *p*-coumaroyl-CoA, starting from photosynthesis, glycolysis, and pentose phosphate pathways (Table 3). However, the next steps in the formation of *p*-HBA have not been clearly determined. Many transcripts appear to encode 3-ketoacyl-CoA thiolases, and one transcript presented homology with the gene encoding 4-hydroxybenzoyl-CoA thioesterase from *Sphingomonas* sp., but the FPKM for this transcript was < 1. Furthermore, the enzymes from the non- β -oxidative pathway were not identified. Although these data suggest that *p*-HBA biosynthesis proceeds via the β -oxidative pathway, it is not possible to ensure which route is actually operating in *Piper gaudichaudianum*.

Thus, to confirm the operant pathway in this species, the transcripts were re-analyzed by comparing them with a second custom database containing the sequences of enzymes from the different pathways involved in benzoate biosynthesis (from plants, fungi, and bacteria). By using this more specific approach, 774 transcripts were annotated (E value $\leq 1e^{-5}$). Of these transcripts, 34 showed significant homology (E value $\leq 1e^{-50}$) and similarity over 80%

Table 3. Enzymes involved in gaudichaudianic acid biosynthesis based on the transcriptome data of *Piper gaudichaudianum*

	Name	K number	EC number	Pathway
E1	3-deoxy-7-phosphoheptulonate synthase	K01626	EC:2.5.1.54	phenylalanine biosynthesis
E2	3-dehydroquinate synthase	K01735	EC:4.2.3.4	phenylalanine biosynthesis
E3	3-dehydroquinate dehydratase / shikimate dehydrogenase	K13832	EC:4.2.1.10 1.1.1.25	phenylalanine biosynthesis
E4	shikimate kinase	K00891	EC:2.7.1.71	phenylalanine biosynthesis
E5	3-phosphoshikimate 1-carboxyvinyltransferase	K00800	EC:2.5.1.19	phenylalanine biosynthesis
E6	chorismate synthase	K01736	EC:4.2.3.5	phenylalanine biosynthesis
E7	chorismate mutase	K01850	EC:5.4.99.5	phenylalanine biosynthesis
E8	arogenate/prephenate dehydratase	K05359	EC:4.2.1.91 4.2.1.51	phenylalanine biosynthesis
E9	bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	K15849	EC:2.6.1.1 2.6.1.78 2.6.1.79	phenylalanine biosynthesis
E10	histidinol-phosphate aminotransferase	K00817	EC:2.6.1.9	phenylalanine biosynthesis
E11	tyrosine aminotransferase	K00815	EC:2.6.1.5	phenylalanine biosynthesis
E12	phenylalanine ammonia-lyase	K10775	EC:4.3.1.24	phenylpropanoid biosynthesis
E13	<i>trans</i> -cinnamate 4-monooxygenase	K00487	EC:1.14.13.11	phenylpropanoid biosynthesis
E14	4-coumarate-CoA ligase	K01904	EC:6.2.1.12	phenylpropanoid biosynthesis
E15	cinnamoyl-CoA hydratase-dehydrogenase	–	–	β-oxidative pathway
E16	3-ketoacyl-CoA thiolase	K07508/K07509/ K07513	EC:2.3.1.16	β-oxidative pathway
E17	4-hydroxybenzoyl-CoA thioesterase	K01075	EC:3.1.2.23	ubiquinone biosynthesis
E18	glyceraldehyde 3-phosphate dehydrogenase	K00134	EC:1.2.1.12	glycolysis / gluconeogenesis
E19	phosphoglycerate kinase	K00927	EC:2.7.2.3	glycolysis / gluconeogenesis
E20	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	K01834	EC:5.4.2.11	glycolysis / gluconeogenesis
E21	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	K15633/K15634	EC:5.4.2.12	glycolysis / gluconeogenesis
E22	enolase	K01689	EC:4.2.1.11	glycolysis / gluconeogenesis
E23	pyruvate kinase	K00873	EC:2.7.1.40	glycolysis / gluconeogenesis
E24	pyruvate dehydrogenase E1 component alpha and beta subunits	K00161/K00162	EC:1.2.4.1	glycolysis / gluconeogenesis
E25	pyruvate decarboxylase	K01568	EC:4.1.1.1	glycolysis / gluconeogenesis
E26	pyruvate dehydrogenase E2 component	K00627	EC:2.3.1.12	glycolysis / gluconeogenesis
E27	acetyl-CoA C-acetyltransferase	K00626	EC:2.3.1.9	terpenoid backbone biosynthesis
E28	hydroxymethylglutaryl-CoA synthase	K01641	EC:2.3.3.10	terpenoid backbone biosynthesis
E29	hydroxymethylglutaryl-CoA reductase (NADPH)	K00021	EC:1.1.1.34	terpenoid backbone biosynthesis
E30	mevalonate kinase	K00869	EC:2.7.1.36	terpenoid backbone biosynthesis
E31	phosphomevalonate kinase	K00938	EC:2.7.4.2	terpenoid backbone biosynthesis
E32	diphosphomevalonate decarboxylase	K01597	EC:4.1.1.33	terpenoid backbone biosynthesis
E33	1-deoxy-D-xylulose-5-phosphate synthase	K01662	EC:2.2.1.7	terpenoid backbone biosynthesis
E34	1-deoxy-D-xylulose-5-phosphate reductoisomerase	K00099	EC:1.1.1.267	terpenoid backbone biosynthesis
E35	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	K00991	EC:2.7.7.60	terpenoid backbone biosynthesis
E36	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	K00919	EC:2.7.1.148	terpenoid backbone biosynthesis
E37	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	K01770	EC:4.6.1.12	terpenoid backbone biosynthesis

Table 3. Enzymes involved in gaudichaudianic acid biosynthesis based on the transcriptome data of *Piper gaudichaudianum* (cont.)

	Name	K number	EC number	Pathway
E38	(<i>E</i>)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	K03526	EC:1.17.7.1 1.17.7.3	terpenoid backbone biosynthesis
E39	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	K03527	EC:1.17.7.4	terpenoid backbone biosynthesis
E40	isopentenyl-diphosphate delta-isomerase	K01823	EC:5.3.3.2	terpenoid backbone biosynthesis
E41	geranyl diphosphate synthase	K14066	EC:2.5.1.1	terpenoid backbone biosynthesis
E42	4-hydroxybenzoate geranyltransferase 2	K13565	EC 2.5.1.93	ubiquinone biosynthesis
E43	tocopherol cyclase	K09834	EC:5.5.1.24	ubiquinone biosynthesis

(Table S2, Supplementary Information). Interestingly, all 34 annotated transcripts are related to *p*-HBA biosynthesis via the β -oxidative pathway, which corroborates this as the main operant pathway in *Piper gaudichaudianum* for *p*-HBA production.

With respect to the terpenoid portion, all the genes codifying the enzymes from both the mevalonate and methylerythritol phosphate pathways were identified in this species, along with those genes encoding isopentenyl-diphosphate isomerase and geranyl-diphosphate synthase (Table 3). This result confirms the simultaneous operation of these two pathways in *Piper gaudichaudianum*, which is in accordance with previous studies.¹⁰ Many transcripts were identified as genes encoding prenyltransferases. Considering only E values $\leq 1e^{-50}$ and FPKM ≥ 1 , 5 transcripts were initially aligned with 4-hydroxybenzoate polyprenyltransferase genes as being involved in ubiquinone biosynthesis. However, these transcripts also present significant homology with 4-hydroxybenzoate geranyltransferase 2 genes (Table S3, Supplementary Information). These enzymes may be related to the prenylation of *p*-HBA that yields the *2H*-pyran moiety after cyclization.

Finally, similar to what was recently observed for *Peperomia obtusifolia*, the transcriptome analysis of the *Piper gaudichaudianum* leaves revealed the presence of tocopherol cyclase.¹² By using the BLASTN²⁰ web tool from NCBI, it was possible to verify that the transcripts from both species present a high degree of similarity (Figure S5, Supplementary Information). Thus, apart from catalyzing the formation of (*S*)-tocopherols, these enzymes may also be responsible for the non-stereoselective cyclization that yields the racemic chromane and chromene moieties. All the enzymes identified that appear to participate in gaudichaudianic acid biosynthesis in *Piper gaudichaudianum* are presented in Table 3, and a proposed scheme for this pathway is shown in Figure 5.

Conclusions

This study presents for the first time the transcriptome analysis of *Piper gaudichaudianum* using a next-generation sequencing approach. Approximately 10% of the transcribed genes identified are somehow involved in secondary metabolism. This approach allowed for the identification of the main active biosynthetic pathways in this species, such as those involved in the formation of terpenoids, phenylpropanoids, flavonoids, and isoquinoline alkaloids. These data corroborate and complement previous studies performed on *Piper gaudichaudianum*. However, the main contribution of this work concerns the biosynthesis of *p*-hydroxybenzoic acid-derived benzopyrans. Despite the advent of sequencing technologies that facilitate the elucidation of plant secondary metabolism, many genes in the benzoic acid derivative biosynthetic network remain to be discovered.⁴² With this study, it was possible to propose that *p*-HBA is produced via the β -oxidative pathway, providing the first insights into its biosynthesis in *Piper* species. Moreover, the transcriptome analysis revealed the presence of prenyltransferases and tocopherol cyclase enzyme genes, which may be responsible for the prenylation and cyclization that yield the *2H*-pyran moiety. These findings are in agreement with those found in other Piperaceae species, such as *Peperomia obtusifolia*.

Experimental

General experimental procedures

Total RNA was extracted using the RNeasy[®] Plant Mini Kit (Qiagen, Hilden, Germany). A NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) was used to determine the concentration and quality of each RNA sample. The quality of the isolated RNA was checked on a 2100 Bioanalyzer (Agilent Technologies,

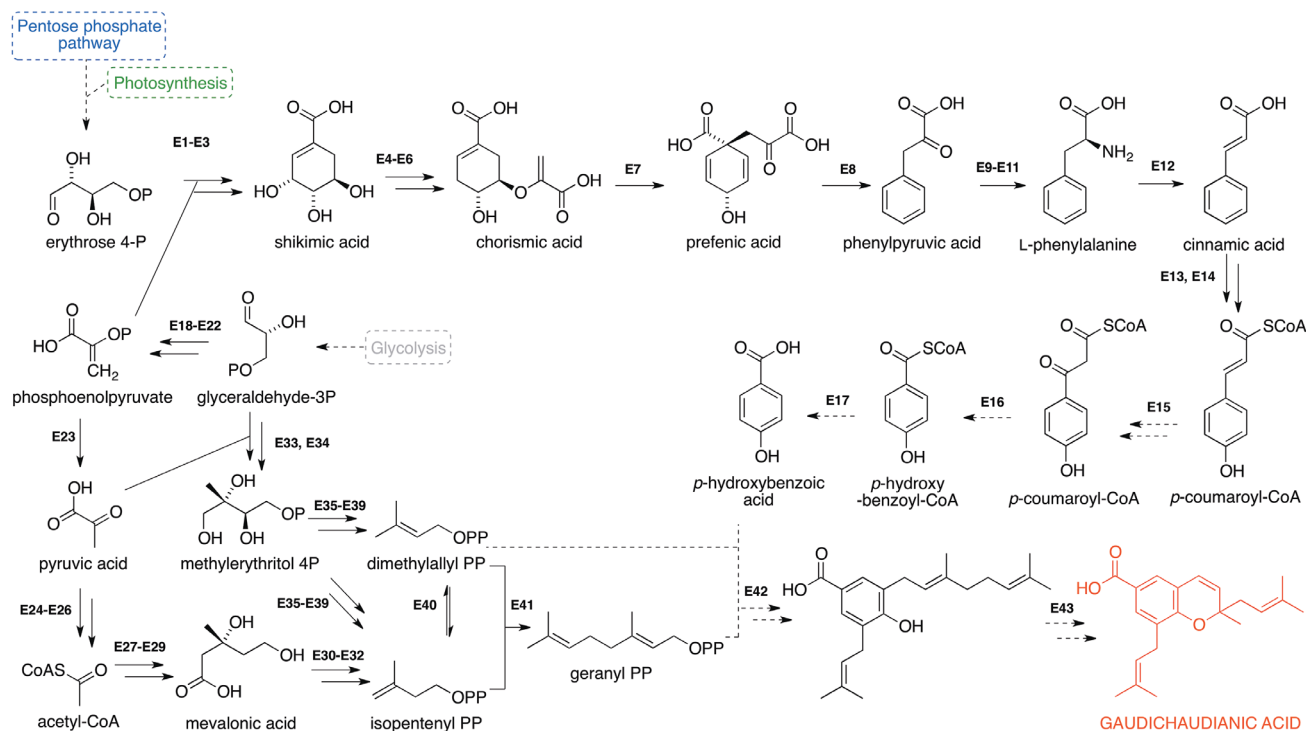


Figure 5. Biosynthesis of gaudichaudianic acid in *Piper gaudichaudianum* including all proteins identified from the transcriptomic study.

Santa Clara, CA, USA). The RNA libraries were prepared using a TruSeq RNA Library Preparation Kit (Illumina, San Diego, CA, USA) and were sequenced using a HiSeq2500 Sequencing System (Illumina, San Diego, CA, USA).

Plant materials

Leaves from young *Piper gaudichaudianum* Kunth specimens, which were cultivated under greenhouse conditions at the Instituto de Química (UNESP, Araraquara, SP, Brazil), were collected. The specimens were identified by Dr Elsie F. Guimarães, and a voucher specimen (Kato-0093) was deposited at the Herbarium of the Instituto de Botânica (São Paulo, SP, Brazil).

RNA extraction, library preparation, and RNA sequencing

Total RNA was extracted in triplicate from a pool of *Piper gaudichaudianum* leaves using the RNeasy® Plant Mini Kit following the manufacturer's instructions. The concentration and quality of the three isolated RNA samples were checked by using the A260/280 and A260/230 ratios from a NanoDrop spectrophotometer. The quality of the samples was also checked in a Bioanalyzer for the presence of intact 28S and 18S bands. Paired-end libraries were prepared using a TruSeq RNA Library Preparation Kit according to the manufacturer's protocol. After that,

the resulting libraries were sequenced using an Illumina HiSeq 2500 device.

De novo transcriptome assembly and annotation

The raw reads obtained after sequencing were quality-filtered using the Trimmomatic software (version 0.33)⁴³ with default parameters to remove the Illumina adapters and low-quality bases. The filtered reads were subjected to a digital normalization algorithm to decrease the sampling variation, discard the redundant data, and remove most of the errors. For links to the digital normalization software, see Supplementary Information section. *De novo* assembly of the filtered clean reads was conducted with the Trinity software,¹⁸ version r20140717, using default parameters (for the assembled sequences, see Supplementary Information section). Fragments *per* kilobase of transcript *per* million fragments mapped (FPKM) values were calculated using the Bowtie2 program.⁴⁴

For annotation, all the assembled transcripts were searched using the BLASTX²⁰ tool with an E value cut-off of $1e^{-5}$ against the following databases: (i) the non-redundant NCBI protein database; (ii) a custom protein database with a total of 948,000 sequences from plant proteins, derived from the RefSeq and UniProt/Swiss-Prot public banks; and (iii) a custom database containing 13,675 enzymes from the different pathways

involved in benzoate biosynthesis (from plants, fungi, and bacteria). This benzoate biosynthesis database included the sequences for chorismate pyruvate lyase, cinnamoyl-CoA hydrates dehydrogenase, 3-ketoacyl-CoA thiolase, 4-hydroxybenzoyl-CoA thioesterase, 1,4-dihydroxy-2-naphthoyl-CoA thioesterase, hydroxycinnamoyl-CoA hydratase lyase, and hydroxybenzaldehyde dehydrogenase, which were derived from the RefSeq and UniProt/Swiss-Prot public banks. The annotation assigned to each transcript was based on the best hit (highest score). The Blast2GO²¹ program was used to assign Gene Ontology (GO) terms to the annotated transcripts according to biological process, molecular function, and cellular component ontologies.

The comparison of the nucleotide sequences to verify their degree of similarity was performed using the BLASTN²⁰ webtool from NCBI.

KEGG pathway mapping assignment

Using the KEGG BlastKOALA²³ annotation web tool, the transcripts with significant hits against the custom plant protein database were also searched against a non-redundant set of KEGG genes from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to assign K numbers. The transcripts with assigned K numbers were mapped using the Search&Color Pathway tool offered by the KEGG database.⁴⁵ This analysis was focused on transcripts with functions that were assigned to a given secondary metabolism biosynthetic pathway.

Supplementary Information

Supplementary information (transcript length distribution, secondary metabolism enzymes (KEGG maps, EC number, K number, FPKM), nucleotide alignment, β -oxidative pathway scheme) is available free of charge at <http://jbcs.s bq.org.br> as PDF file.

All links to the digital normalization software are available electronically through <http://ged.msu.edu/papers/2012-diginorm/>.

The assembled sequences file is available free of charge at <http://jbcs.s bq.org.br>.

Acknowledgments

The authors are grateful to the São Paulo Research Foundation (FAPESP, grant numbers 2013/07600-3 (CEPID-CIBFar), 2014/25222-9, 2014/50316-7 and 2015/07089-2) and to the National Council for Scientific and Technological Development (CNPq) for the research fellowships to M. F. and M. J. K.; A. N. L. B. thanks

FAPESP for the provision of a postdoctoral fellowship (grant number 2011/01003-8).

References

1. Kato, M. J.; Furlan, M.; *Pure Appl. Chem.* **2007**, *79*, 529.
2. López, S. N.; Lopes, A. A.; Batista Jr., J. M.; Flausino Jr., O.; Bolzani, V. S.; Kato, M. J.; Furlan, M.; *Bioresour. Technol.* **2010**, *101*, 4251.
3. Van Otterlo, W. A. L.; Ngidi, E. L.; Kuzvidza, S.; Morgans, G. L.; Moleele, S. S.; de Koning, C. B.; *Tetrahedron* **2005**, *61*, 9996.
4. Nicolaou, K. C.; Pfefferkorn, J. A.; Roecker, A. J.; Cao, G. Q.; Barluenga, S.; Mitchell, H. J.; *J. Am. Chem. Soc.* **2000**, *122*, 9939.
5. Tang, Y.; Wei, J.; Zhong, W.; Liu, X.; *Heteroat. Chem.* **2010**, *21*, 423.
6. Batista Jr., J. M.; Batista, A. N. L.; Rinaldo, D.; Vilegas, W.; Ambrósio, D. L.; Cicarelli, R. M. B.; Bolzani, V. S.; Kato, M. J.; Nafie, L. A.; López, S. N.; Furlan, M.; *J. Nat. Prod.* **2011**, *74*, 1154.
7. Mota, J. S.; Leite, A. C.; Batista Jr., J. M.; 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.
8. Batista Jr., J. M.; 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.
9. Batista Jr., J. M.; 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.
10. Lopes, A. A.; Baldoqui, D. C.; López, S. N.; Kato, M. J.; Bolzani, V. S.; Furlan, M.; *Phytochemistry* **2007**, *68*, 2053.
11. Salazar, K. J. M.; Paredes, G. E. D.; Llluncor, L. R.; Young, M. C. M.; Kato, M. J.; *Phytochemistry* **2005**, *66*, 573.
12. Batista, A. N. L.; Santos-Pinto, J. R.; Batista Jr., J. M.; Souza-Moreira, T.; Santoni, M. M.; Zanelli, C. F.; Kato, M. J.; López, S. N.; Palma, M. S.; Furlan, M.; *J. Nat. Prod.* **2017**, *80*, 1275.
13. Batista Jr., J. M.; Lopes, A. A.; Ambrósio, D. L.; Regasini, L. O.; Kato, M. J.; Bolzani, V. S.; Cicarelli, R. M. B.; Furlan, M.; *Biol. Pharm. Bull.* **2008**, *31*, 538.
14. Gaia, A. M.; Yamaguchi, L. F.; Jeffrey, C. S.; Kato, M. J.; *Phytochemistry* **2014**, *106*, 86.
15. Wang, Z.; Gerstein, M.; Snyder, M.; *Nat. Rev. Genet.* **2009**, *10*, 57.
16. Góngora-Castillo, E.; Buell, C. R.; *Nat. Prod. Rep.* **2013**, *30*, 490.
17. Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A.; *Bioinformatics* **2014**, *30*, 614.
18. 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.
19. Hu, L.; Hao, C.; Fan, R.; Wu, B.; Tan, L.; Wu, H.; *PLoS One* **2015**, *10*, e0129822.
20. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J.; *J. Mol. Biol.* **1990**, *215*, 403.
21. Conesa, A.; Götz, S.; *Int. J. Plant Genomics* **2008**, 19832.
22. Ong, W. D.; Voo, L.-Y. C.; Kumar, V. S.; *PLoS One* **2012**, *7*, e46937.
23. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M.; *Nucleic Acids Res.* **2016**, *44*, D457.
24. Péres, V. F.; Saffi, J.; Melecchi, M. I.; Abad, F. C.; Jacques, R. A.; Martinez, M. M.; Oliveira, E. C.; Caramão, E. B.; *J. Chromatogr. A* **2006**, *1105*, 115.
25. Péres, V. F.; Moura, D. J.; Sperotto, A. R.; Damasceno, F. C.; Caramão, E. B.; Zini, C. A.; Saffi, J.; *Food Chem Toxicol.* **2009**, *47*, 2389.
26. Sperotto, A. R.; Moura, D. J.; Péres, V. F.; Damasceno, F. C.; Caramão, E. B.; Henriques, J. A.; Saffi, J.; *Food Chem. Toxicol.* **2013**, *57*, 57.
27. Yamaguchi, L. F.; Freitas, G. C.; Yoshida, N. C.; Silva, R. A.; Gaia, A. M.; Silva, A. M.; Scotti, M. T.; Emerenciano, V. P.; Guimarães, E. F.; Floh, E. I. S.; Colombo, C. A.; Siqueira, W. J.; Kato, M. J.; *J. Braz. Chem. Soc.* **2011**, *22*, 2371.
28. Lago, J. H. G.; Ramos, C. S.; Casanova, D. C. C.; Morandim, A. A.; Bergamo, D. C. B.; Cavalheiro, A. J.; Bolzani, V. S.; Furlan, M.; Guimarães, E. F.; Young, M. C. M.; Kato, M. J.; *J. Nat. Prod.* **2004**, *67*, 1783.
29. Rapado, L. N.; Freitas, G. C.; Polpo, A.; Rojas-Cardozo, M.; Rincón, J. V.; Scotti, M. T.; Kato, M. J.; Nakano, E.; Yamaguchi, L. F.; *Molecules* **2014**, *19*, 5205.
30. Bentley, K. W.; *Nat. Prod. Rep.* **2006**, *23*, 444.
31. Sircar, D.; Mitra A.; *J. Plant. Physiol.* **2008**, *165*, 407.
32. Mustafa, N. R.; Verpoorte, R.; *Planta* **2005**, 222, 1.
33. Block, A.; Widhalm, J. R.; Fatihi, A.; Cahoon, R. E.; Wamboldt, Y.; Elowsky, C.; Mackenzie, S. A.; Cahoon, E. B.; Chapple, C.; Dudareva, N.; Basset, G. J.; *Plant Cell.* **2014**, *26*, 1938.
34. Kawamukai, M.; *Biotechnol. Appl. Biochem.* **2009**, *53*, 217.
35. Meganathan, R.; *FEMS Microbiol. Lett.* **2001**, *203*, 131.
36. Van Moerkercke, A.; Schauvinhold, I.; Pichersky, E.; Haring, M. A.; Schuurink, R. C.; *Plant J.* **2009**, *60*, 292.
37. Loscher, R.; Heide, L.; *Plant Physiol.* **1994**, *106*, 271.
38. Podstolski, A.; Havkin-Frenkel, D.; Malinowski, J.; Blount, J. W.; Kourteva, G.; Dixon, R. A.; *Phytochemistry* **2002**, *61*, 611.
39. Qualley, A. V.; Widhalm, J. R.; Adebessin, F.; Kish, C. M.; Dudareva, N.; *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16383.
40. Widhalm, J. R.; Ducluzeau, A.-L.; Buller, N. E.; Elowsky, C. G.; Olsen, L. J.; Basset, G. J. C.; *Plant J.* **2012**, *71*, 205.
41. McQualter, R. B.; Chong, B. F.; Meyer, K.; Van Dyk, D. E.; O'Shea, M. G.; Walton, N. J.; Viitanen, P. V.; Brumbley, S. M.; *Plant Biotechnol. J.* **2005**, *3*, 29.
42. Widhalm, J. R.; Dudareva, N.; *Mol. Plant* **2015**, *8*, 83.
43. Bolger, A. M.; Lohse, M.; Usadel, B.; *Bioinformatics* **2014**, *30*, 2114.
44. Langmead, B.; Salzberg, S.; *Nat. Methods* **2012**, *9*, 357.
45. http://www.genome.jp/kegg/tool/map_pathway2.html, accessed in December 2017.

Submitted: October 20, 2017

Published online: December 20, 2017