# Coupling Deep Transcriptome Analysis with Untargeted Metabolic Profiling in *Ophiorrhiza pumila* to Further the Understanding of the Biosynthesis of the Anti-Cancer Alkaloid Camptothecin and Anthraquinones

Mami Yamazaki[1,9], Keiichi Mochida[2,3,4,9], Takashi Asano[1,10], Ryo Nakabayashi[2], Motoaki Chiba[1], Nirin Udomson[1], Yasuyo Yamazaki[5], Dayan B. Goodenowe[5], Ushio Sankawa[6,11], Takuhiro Yoshida[2], Atsushi Toyoda[7,12], Yasushi Totoki[7,13], Yoshiyuki Sakaki[7,14], Elsa Góngora-Castillo[8], C. Robin Buell[8], Tetsuya Sakurai[2] and Kazuki Saito[1,2,*]

[1]Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chiba, 260-8675 Japan
[2]RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan
[3]RIKEN Biomass Engineering Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan
[4]Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka-cho, Totsuka-ku, Yokohama, Kanagawa, 244-0813 Japan
[5]Phenomenome Discoveries Inc., Saskatoon, SK, Canada
[6]International Research Center for Traditional Medicine, Toyama Prefecture, Toyama, 939-8224 Japan
[7]RIKEN Genome Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan
[8]Department of Plant Biology, Michigan State University, 612 Wilson Rd, Plant Biology Laboratories, East Lansing, MI 48824-1312, USA
[9]These authors contributed equally to this work.
[10]Present address: School of Pharmacy, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba, Iwate, 028-3694 Japan.
[11]Present address: The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033 Japan.
[12]Present address: National Institute of Genetics, Yata 1111, Mishima, Shizuoka, 411-8540 Japan.
[13]Present address: National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.
[14]Present address: Toyohashi University of Technology, 1-1, Hibarigaoka, Tenpaku-cho, Toyohashi, Aichi, 441-8580 Japan.
*Corresponding author: E-mail, ksaito@faculty.chiba-u.jp; Fax, +81-43-226-2932.

The Rubiaceae species, *Ophiorrhiza pumila*, accumulates camptothecin, an anti-cancer alkaloid with a potent DNA topoisomerase I inhibitory activity, as well as anthraquinones that are derived from the combination of the isochorismate and hemiterpenoid pathways. The biosynthesis of these secondary products is active in *O. pumila* hairy roots yet very low in cell suspension culture. Deep transcriptome analysis was conducted in *O. pumila* hairy roots and cell suspension cultures using the Illumina platform, yielding a total of 2 Gb of sequence for each sample. We generated a hybrid transcriptome assembly of *O. pumila* using the Illumina-derived short read sequences and conventional Sanger-derived expressed sequence tag clones derived from a full-length cDNA library constructed using RNA from hairy roots. Among 35,608 non-redundant unigenes, 3,649 were preferentially expressed in hairy roots compared with cell suspension culture. Candidate genes involved in the biosynthetic pathway for the monoterpenoid indole alkaloid camptothecin were identified; specifically, genes involved in post-strictosamide biosynthetic events and genes involved in the biosynthesis of anthraquinones and chlorogenic acid. Untargeted metabolomic analysis by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) indicated that most of the proposed intermediates in the camptothecin biosynthetic pathway accumulated in hairy roots in a preferential manner compared with cell suspension culture. In addition, a number of anthraquinones and chlorogenic acid preferentially accumulated in hairy roots compared with cell suspension culture. These results suggest that deep transcriptome and metabolome data sets can facilitate the identification of genes and intermediates involved in the biosynthesis of secondary products including camptothecin in *O. pumila*.

liquid chromatography–mass spectrometry, MEP; 2-C-methyl-D-erythritol 4-phosphate.

The nucleotide sequences reported in this paper have been submitted to DDBJ under the accession numbers, HX703880–HX739052, for ESTs; and DDBJ Sequence Read Archive under the accession number, DRA000930, for the Illumina reads.

## Introduction

In the post-genomic era, integrated analysis of comprehensive gene expression (transcriptome) and metabolic profiling (metabolomics) has been successful in plant functional genomics (Saito et al. 2008, Saito and Matsuda 2010). In particular, a number of examples in *Arabidopsis thaliana* have reported the identification and/or the prediction of gene function employing a coupled approach using transcriptome and metabolome data sets (Hirai et al. 2005, Tohge et al. 2005, Hirai et al. 2007, Yonekura-Sakakibara et al. 2007, Yonekura-Sakakibara et al. 2008, Okazaki et al. 2009, Matsuda et al. 2010, Yonekura-Sakakibara et al. 2012). In *A. thaliana*, these successes are due in large part to the availability of genome sequence data, expression profiles and genetic resources such as tagged mutants and natural variants that enable functional genomics research. The application of integrated 'omics' analyses to medicinal plants, however, is not necessarily straightforward, as typically few, if any, sequence or expression profiling data are available (Yonekura-Sakakibara and Saito 2009). Nevertheless, the recent technological advances in DNA sequencing, the so-called 'next-generation sequencing', allow holistic profiling of RNA expression (Wang et al. 2009, Ozsolak and Milos 2011) in non-model plant species in which limited molecular genetics studies have been performed. Deep transcriptome technology, also known as RNA sequencing (RNA-seq), provides whole-transcriptome expression profiles of selected plant tissues or cells, thereby permitting the integrated analysis of transcriptomics and metabolomics in any plant species.

The anti-cancer alkaloid, camptothecin (compound **7** in **Fig. 1**), was identified first in the Chinese tree *Camptotheca acuminata* (Nyssaceae) during an extensive screening of natural anti-cancer compounds (Wall et al. 1966). This alkaloid inhibits DNA topoisomerase I and has potent anti-tumor activity (Hsiang et al. 1985). In the clinical field, semi-synthetic water-soluble camptothecin derivatives are used throughout the world as chemotherapy agents for cancers of the lung, cervix,
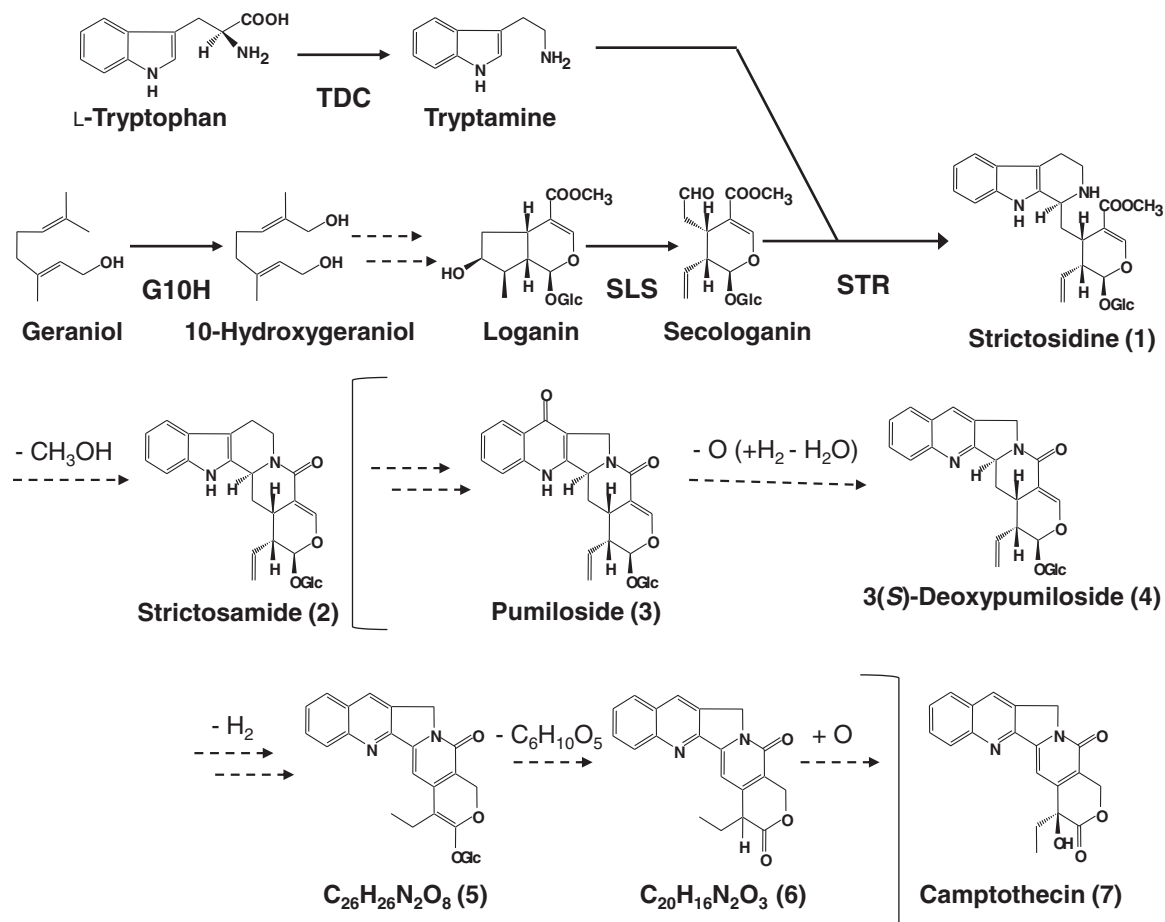


**Fig. 1** Possible biosynthetic pathway of camptothecin and related alkaloids.

ovary, colon and other organs (Sirikantaramas et al. 2007). Though the skeleton structure is a pentacyclic quinoline-type alkaloid, camptothecin is synthesized via the monoterpenopid indole alkaloid pathway (Hutchinson et al. 1979, Yamazaki et al. 2003a, Yamazaki et al. 2004) (**Fig. 1**).

The genus *Ophiorrhiza* (Rubiaceae) encompasses approximately 150 species distributed throughout tropical and subtropical Asia (Darwin 1976), with some species producing camptothecin and related indole alkaloids (Aimi et al. 1989, Arbain et al. 1993, Kitajima et al. 2005, Viraporn et al. 2011). *Ophiorrhiza pumila* specimens in Japan accumulate camptothecin and its potential biosynthetic intermediates (Aimi et al. 1989, Aimi et al. 1990). *Ophiorrhiza pumila* also produces anthraquinones, flavonoids and chlorogenic acid, in addition to indole alkaloids (Kitajima et al. 1998, Yamazaki et al. 2003b). Several genes involved in the biosynthesis of camptothecin have been characterized (Yamazaki et al. 2003a), and possible intermediates in camptothecin biosynthesis have been proposed (Asano et al. 2012). Nevertheless, a large part of the information on the genes and their encoded enzymes for camptothecin biosynthesis is as yet unknown.

We established an *O. pumila* hairy root culture (Saito et al. 2001) which produces camptothecin and anthraquinones, and an *O. pumila* cell suspension culture derived from hairy roots (Asano et al. 2012) which does not accumulate these secondary metabolites. These two cultured cells provide contrasting sources for integrated and differential analysis of transcriptome and metabolome data to facilitate the prediction of candidate genes involved in the biosynthesis of these secondary metabolites. In the present study, we utilized deep transcriptomics combined with untargeted metabolic profiling by in-fusion Fourier transform-ion cyclotron resonance-mass spectrometry (FT-ICR-MS) using a hairy root culture and a cell suspension culture of *O. pumila* to identify candidate genes associated with secondary metabolite production. Through our combined analyses, we identified differential transcripts and metabolites in the two tissues that are presumed to be associated with the biosynthesis of secondary products including camptothecin. These data sets are useful resources for further studies involving intermediary metabolites.

## Results

### Assembly of deep transcriptome data from *O. pumila*

Deep transcriptome analysis was performed using RNA from hairy roots and cell suspension culture, exhibiting differential accumulation of secondary metabolites, using the Illumina HiSeq 2000 platform. Over 20 million paired-end reads for each sample representing >2 Gb were generated, and a transcriptome was generated with these two data sets using the Oases transcriptome assembler (Schulz et al. 2012). A full-length cDNA library was constructed from *O. pumila* hairy

roots, and 35,208 expressed sequence tags (ESTs) from 19,400 clones were sequenced using Sanger dideoxy sequencing. The Sanger-derived EST data were used in a hybrid assembly with the Oases-derived transcript assemblies, yielding 35,608 non-redundant unigenes that represent the *O. pumila* transcriptome. To estimate expression abundances in the two tissues, reads from the hairy root and the cell suspension culture RNA-seq data sets were mapped to the *O. pumila* unigene set using Bowtie (Langmead et al. 2009); in total, 87–91% of reads could be mapped to the assembled non-redundant unigenes (**Supplementary Fig. S1**).

### Overview of differentially expressed genes between hairy roots and cell suspension culture

Expression levels in the hairy roots or cell suspension culture libraries were evaluated using Cufflinks (Trapnell et al. 2010) (**Supplementary Fig. S1**). Differential expression was determined using a false discovery rate (FDR) threshold of corrected $P$-value $<0.05$ (**Fig. 2A**), and a heatmap diagram sorted by the magnitude of fold change (log2) is shown in **Fig. 2B**. Using differential expression analyses, a total of 3,649 unigenes were up-regulated in hairy roots compared with cell suspension culture, whereas 1,777 unigenes were down-regulated in hairy roots compared with cell suspension culture (**Supplementary Table S1**).

Gene ontology (GO) enrichment analysis was performed with the 3,649 unigenes up-regulated in hairy roots compared with cell suspension cells (**Supplementary Fig. S2**, **Supplementary Table S2**). The GO terms 'oxidation–reduction process' (GO: 0055114) and 'organic substance transport' (GO: 0071702) were the most significantly enriched, followed by 'developmental maturation' (GO: 0021700) and 'transition metal ion transport' (GO: 0000041). Other significantly enriched GO terms in the set of 3,649 up-regulated unigenes included various 'transport' ontologies, 'response to oxidative stress', 'phenylpropanoid metabolic process' and those related to developmental processes including the terms 'cell maturation' and 'root hair cell differentiation'. The GO analysis suggested that the genes related to secondary metabolism and hairy root development by *Agrobacterium rhizogenes* infection were particularly enriched.

### Specific genes involved in the biosynthesis of secondary metabolites in hairy roots

*Camptothecin biosynthetic genes.* For the biosynthesis of camptothecin in *O. pumila*, several genes encoding enzymes involved in the pathway, as well as DNA topoisomerase I which is the target site of camptothecin toxicity, have been cloned (Yamazaki et al. 2003a, Sirikantaramas et al. 2008). The preferential expression of these known camptothecin biosynthetic pathway genes was examined in the transcriptome data set (**Table 1**). Three unigenes encoding tryptophan decarboxylase and one unigene for strictosidine synthase, both involved in the early steps of the biosynthesis of
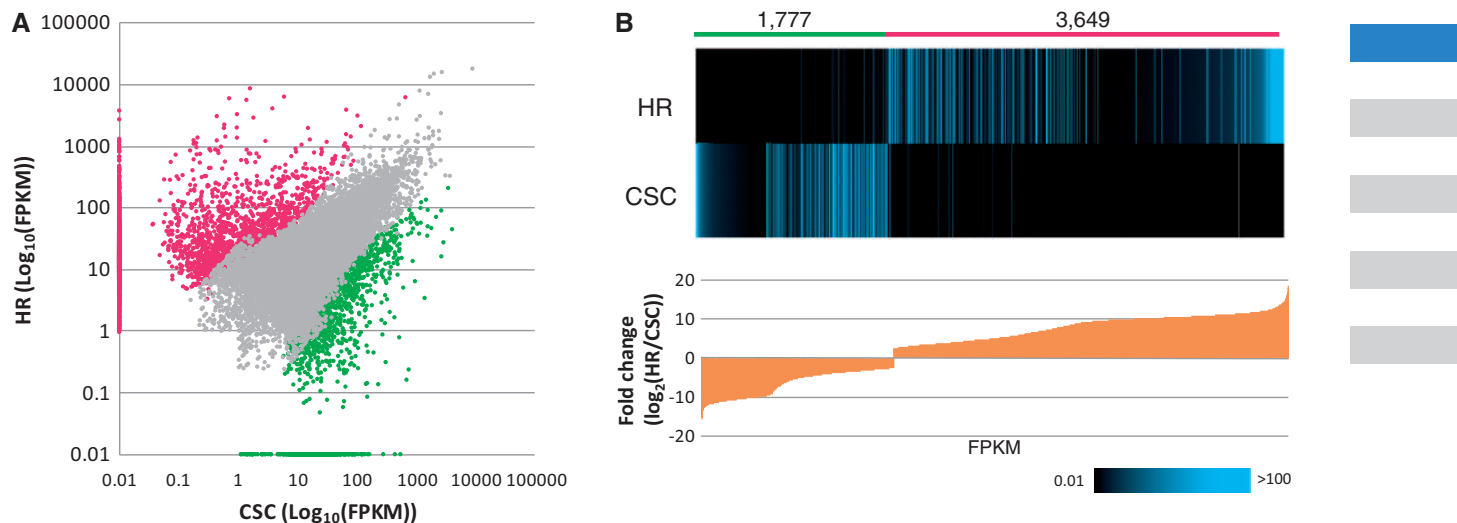
**Fig. 2** Deep transcriptome analysis of hairy roots and cell suspension cultures of *O. pumila*. (A) Expression profile of all unigenes plotted by log10 FPKM in hairy roots (HR) (*y*-axis) and cell suspension culture (CSC) (*x*-axis). The unigenes differentially expressed with a threshold of an FDR-corrected *P*-value <0.05 are shown in color: red, up-regulated inHR; green, up-regulated in CSC. (B) The differentially expressed unigenes are shown in a heatmap diagram sorted by the magnitude of fold change (log2). The 3,649 and 1,777 unigenes from HR and CSC are preferentially expressed in the respective samples.

**Table 1** Expression of previously characterized genes related to camptothecin biosynthesis and self-resistance of *O. pumila*

| | FPKM | | BLAST hit of the NCBI nr data set (BLASTx) | | |
|---|---|---|---|---|---|
| | CSC | HR | Accession No. | Description | E-value |
| Contig7117 | 0.05 | 79.06 | AB086168 | Tryptophan decarboxylase (*Ophiorrhiza pumila*) | 1.00E-180 |
| Contig8774 | 0.22 | 544.06 | AB086168 | Tryptophan decarboxylase (*Ophiorrhiza pumila*) | 6.00E-88 |
| 21504_iso_9 | 0.00 | 137.93 | AB086168 | Tryptophan decarboxylase (*Ophiorrhiza pumila*) | 1.00E-180 |
| Contig8730 | 0.06 | 556.54 | AB060341 | Strictosidine synthase (*Ophiorrhiza pumila*) | 0 |
| Contig9209 | 99.91 | 116.06 | AB086169 | NADPH-cytochrome P-450 reductase (*Ophiorrhiza pumila*) | 0 |
| Contig9971 | 51.37 | 38.10 | AB372508 | Topoisomerase I (*Ophiorrhiza pumila*) | 0 |
| 1191_iso_2 | 56.27 | 37.68 | AB372508 | Topoisomerase I (*Ophiorrhiza pumila*) | 0 |

Expression was measured as fragments per kilobase of transcript per million fragments mapped (FPKM). Functional annotation was assigned from identical BLAST hit's description of the NCBI nr data set. The expression levels of unigenes are represented with a black–blue color code and FPKM values.
CSC, cell suspension culture; HR, hairy roots.

monoterpenoid indole alkaloids, were highly differentially expressed in hairy roots compared with cell suspension cells. However, unigenes of two enzymes which are not specifically involved in the pathway were equally expressed in hairy roots and cell suspension culture. These are NADPH-cytochrome P450 reductase, which transfers an electron from NADPH to Cyt P450 non-specifically, and DNA topoisomerase I, which is involved in DNA replication and acts as the target site of camptothecin toxicity.

*Anthraquinone biosynthetic genes.* Anthraquinones are a major group of secondary products found in the plants of the Rubiaceae family including *O. pumila* (Kitajima et al. 1998, Chan et al. 2005). Anthraquinones in Rubiaceae species have been shown to be synthesized by a combination of the isochorismate

and plastidic hemiterpenoid 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathways (Han et al. 2001, Han et al. 2002). The anthraquinone skeleton is formed by coupling of a 1,4-dihydroxy-2-naphthoyl derivative with dimethylallyl diphosphate (**Fig. 3**). Thus, we examined the expression of unigenes encoding enzymes for these two pathways based on fragments per kilobase of transcript per million fragments mapped (FPKM). All the unigenes encoding the enzymes involved in the formation of 1,4-dihydroxy-2-naphthoyl-CoA from chorismate were significantly highly expressed in the hairy roots compared with cell suspension culture. On the other hand, regarding the plastidic hemiterpenoid MEP pathway, only the unigenes encoding four enzymes were highly differentially expressed in the hairy roots compared with cell suspension culture (**Fig. 3**). The unigenes that encode

| MenF | CSC | HR |
|---|---|---|
| Contig6581 | 1.39 | 137.54 |
| Contig6582 | 0.00 | 49.84 |

| MenD, MenH, MenC | CSC | HR |
|---|---|---|
| Contig7102 | 2.82 | 79.55 |
| Contig7191 | 4.01 | 91.65 |

| MenE | CSC | HR |
|---|---|---|
| OPN011_F01 | 2.04 | 234.18 |
| OPN014_O14 | 0.67 | 96.23 |
| 21282_iso_5 | 0.63 | 130.03 |
| Contig7101 | 0.53 | 170.57 |

| MenB | CSC | HR |
|---|---|---|
| Contig7051 | 5.00 | 447.22 |
| Contig7891 | 4.22 | 460.08 |
| Contig7958 | 0.90 | 115.35 |

| CLA1, DXPS3 | CSC | HR |
|---|---|---|
| Contig11115 | 0.00 | 41.46 |
| Contig2771 | 8.18 | 159.13 |
| Contig6147 | 11.04 | 8.39 |
| Contig8042 | 0.08 | 53.88 |
| Contig9037 | 10.15 | 10.41 |
| OPN050_K04 | 0.52 | 2.52 |
| 28410_iso_1 | 10.93 | 8.21 |

| DXR | CSC | HR |
|---|---|---|
| Contig10335 | 90.06 | 113.52 |
| Contig420 | 15.12 | 14.57 |
| Contig421 | 19.44 | 24.33 |
| Contig422 | 33.20 | 46.14 |
| Contig8340 | 180.60 | 212.63 |
| 661_iso_6 | 30.72 | 41.50 |
| 661_iso_9 | 46.02 | 59.34 |
| 7932_iso_1 | 387.90 | 434.10 |

| ISPD | CSC | HR |
|---|---|---|
| Contig5108 | 27.051 | 27.4 |

| CDPMEK | CSC | HR |
|---|---|---|
| Contig3032 | 17.38 | 58.29 |
| OPN005_J15 | 4.606 | 12.85 |

| ISPF | CSC | HR |
|---|---|---|
| Contig4110 | 112.9 | 175.7 |

| HDS | CSC | HR |
|---|---|---|
| Contig2436 | 38.92 | 112.3 |
| Contig8588 | 38.95 | 112.1 |

| HDR | CSC | HR |
|---|---|---|
| Contig10924 | 18.26 | 51.62 |
| Contig582 | 20.63 | 61.32 |
| Contig583 | 25.22 | 70.75 |
| 935_iso_1 | 19.22 | 59.79 |
| 935_iso_4 | 17.6 | 52.37 |
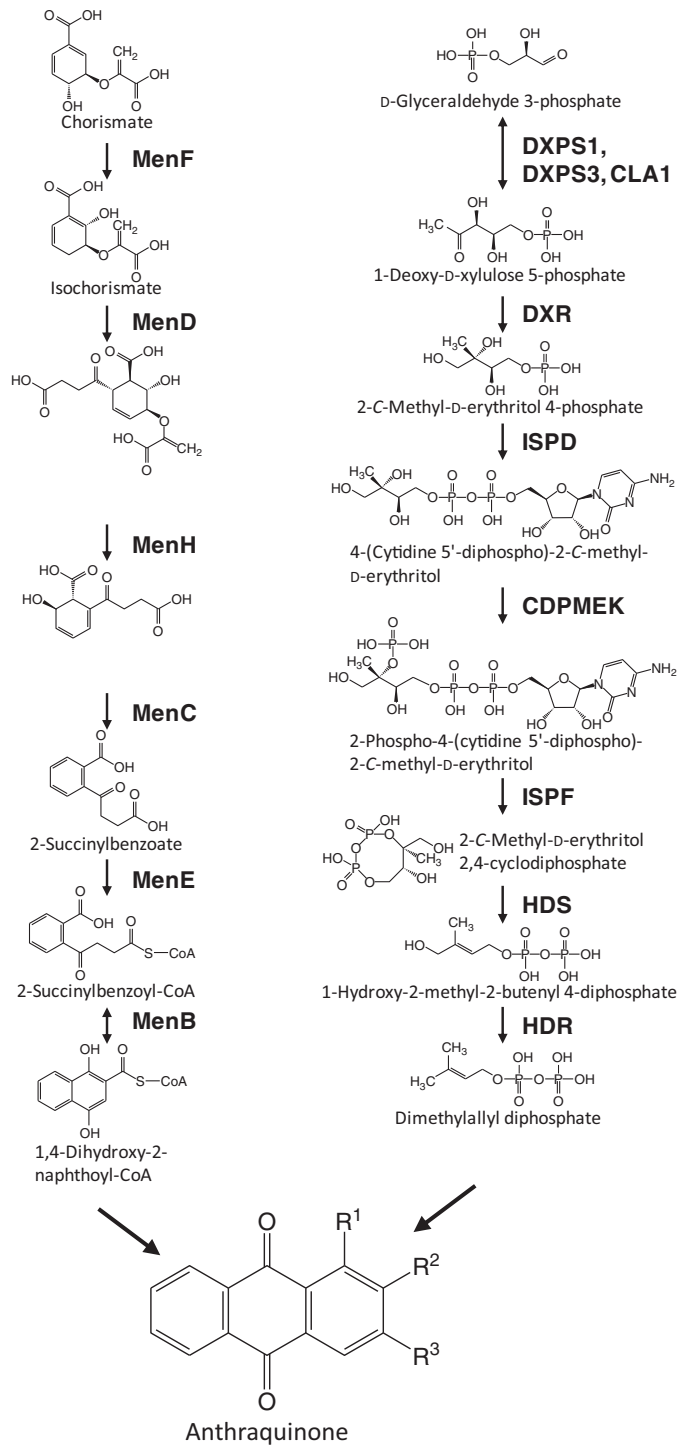
Pathway intermediates (left to right): Chorismate → (MenF) → Isochorismate → (MenD) → (MenH) → (MenC) → 2-Succinylbenzoate → (MenE) → 2-Succinylbenzoyl-CoA → (MenB) → 1,4-Dihydroxy-2-naphthoyl-CoA.

D-Glyceraldehyde 3-phosphate → (DXPS1, DXPS3, CLA1) → 1-Deoxy-D-xylulose 5-phosphate → (DXR) → 2-C-Methyl-D-erythritol 4-phosphate → (ISPD) → 4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol → (CDPMEK) → 2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol → (ISPF) → 2-C-Methyl-D-erythritol 2,4-cyclodiphosphate → (HDS) → 1-Hydroxy-2-methyl-2-butenyl 4-diphosphate → (HDR) → Dimethylallyl diphosphate.

Anthraquinone (with substituents R$^1$, R$^2$, R$^3$).

**Fig. 3** Plausible biosynthetic pathway and expression of unigenes involved in the biosynthesis of anthraquinones in *O. pumila*. The expression levels of unigenes encoding the enzymes of each step are shown. Separated proteins of MenD, MenH and MenC in bacteria are known to be fused to a single protein in Arabidopsis (Gross et al. 2006). HR, hairy roots; CSC, cell suspension culture; MenF, menaquinone-specific isochorismate synthase; MenD, 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase; MenH, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase; MenC, *o*-succinylbenzoate synthase; MenE, *o*-succinylbenzoic acid-CoA ligase; MenB, naphthoate synthase; DXPS, 1-deoxy-D-xylulose-5-phosphate synthase; CLA, cloroplastos alterados; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ISPD, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CDPMEK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; ISPF, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase. The expression levels of unigenes are represented with a black–blue color code and FPKM values.

**Table 2** Expression abundances of unigenes potentially involved in the biosynthesis of chlorogenic acid

| | FPKM | | BLAST hit of the TAIR10 proteome dataset (BLASTx) | | |
|---|---|---|---|---|---|
| | CSC | HR | Description | | E-value |
| Contig119240:−1571 | 0.0 | 22.2 | HCT | AT5G48930.1 | 2.00E-57 |
| Contig82420:−321 | 0.0 | 28.9 | HCT | AT5G48930.1 | 2.00E-11 |
| Contig83070:−2120 | 0.5 | 91.0 | HCT | AT5G48930.1 | 1.00E-158 |
| opa_locus_4548_iso_1 | 14.8 | 99.0 | HCT | AT5G48930.1 | 0 |
| opa_locus_56584_iso_1 | 0.0 | 9.5 | HCT | AT5G48930.1 | 2.00E-14 |
| Contig11650 | 34 | 203 | CYP98A3 | AT2G40890.1 | 0 |
| Contig8316 | 19.5 | 150 | CYP98A3 | AT2G40890.1 | 4.00E-82 |
| Contig9862 | 3.96 | 95.5 | CYP98A3 | AT2G40890.1 | 0 |
| OPN002_A17.r | 2.18 | 18.2 | CYP98A3 | AT2G40890.1 | 1.00E-47 |
| OPN018_J04.r | 0.38 | 3.65 | CYP98A3 | AT2G40890.1 | 3.00E-33 |
| OPN044_K16.r | 9.06 | 41.9 | CYP98A3 | AT2G40890.1 | 2.00E-18 |

Expression was measured as fragments per kilobase pair of transcript per million fragments mapped (FPKM). Functional annotation was deduced from BLAST best homolog's description of the TAIR10 proteome. The expression levels of unigenes are represented with a black–blue color code and FPKM values.
HR, hairy roots; CSC, cell suspension culture.

1-deoxy-D-xylulose-5-phosphate synthase (DXPS; EC 2.2.1.7) (Contig11115, Contig2771 and Contig8042), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CDPMEK; EC 2.7.1.148), (*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS; EC 1.17.7.1) and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR; EC 1.17.1.2) were highly differentially expressed in the hairy roots, suggesting that these steps may be rate limiting in the formation of dimethylallyl diphosphate leading to anthraquinones.

*Chlorogenic acid biosynthetic genes.* In addition to camptothecin and anthraquinones, chlorogenic acid (3-O-caffeoyl quinic acid) (**21**) preferentially accumulates in *O. pumila* hairy roots (see below) (Yamazaki et al. 2003b). Chlorogenic acid is synthesized by a two-step reaction from *p*-coumaroyl-CoA; namely *p*-coumaroyl transfer to quinic acid by shikimate *O*-hydroxycinnamoyltransferase (HCT; EC 2.3.1.133) and subsequent hydroxylation of the 3′ position of coumaroylquinate by coumaroylquinate(coumaroylshikimate) 3′-monooxygenase (CYP98A3, C3′H; EC 1.14.13.36), or alternatively by hydroxylation of the 3′ postion of *p*- coumaroyl-CoA first and transfer to quinic acid (Koshiro et al. 2007) (http://www.kegg.jp/kegg-bin/show_pathway?map00941, http://pmn.plantcyc.org/PLANT/NEW-IMAGE?type = PATHWAY&object = PWY-6039) (Kanehisa et al. 2010, Chae et al. 2012). Unigenes with sequence similarity to Arabidopsis HCT and C3′H, were highly differentially expressed in hairy roots compared with cell suspension culture (**Table 2**). These unigenes presumably encode HCT and the C3′H involved in the formation of chlorogenic acid in *O. pumila*.

Collectively, these results indicate that differential expression analysis can assist in the delineation of genes involved in the biosynthesis of secondary products, including camptothecin, anthraquinones and chlorogenic acid in *O. pumila* hairy roots.

## Untargeted metabolome analysis by in-fusion FT-ICR-MS and differential accumulation of secondary metabolites

We previously investigated the metabolic profiles of hairy roots and cell suspension culture using liquid chromatography–mass spectrometry (LC-MS) analysis (Yamazaki et al. 2003b). In the present study, we adopted an untargeted metabolomic analysis using in-fusion FT-ICR-MS (**Table 3**; **Supplementary Table S3**) (Hirai et al. 2005, Tohge et al. 2005). Annotation of detected peaks in the analyses was based on comparison of accurate *m/z* values observed with the theoretical values deduced from the elemental composition within a certain tolerance range. The determined elemental composition was used for putative identification of compounds by retrieving data within plant metabolite databases, KNApSAcK Core (Afendi et al. 2012) and A Dictionary of Natural Products (Hocking 1997). A total of 599 peaks were differentially accumulated, ranging from 0.022- to 303-fold (hairy root/cell suspension) (*P*-value <0.01, *t*-test) (**Supplementary Table S3**). Among them, 327 peaks were preferentially accumulated in hairy roots, and 272 peaks were preferentially accumulated in the cell suspension culture. Most of the proposed intermediates in the pathway of camptothecin biosynthesis accumulated in a hairy root-specific manner. In addition to camptothecin pathway intermediates, a number of anthraquinone-related compounds were highly differentially accumulated in hairy roots compared with the cell suspension culture. Chlorogenic acid and some other unidentified metabolites, which have been previously detected in hairy roots by LC-MS (Yamazaki et al. 2003b), were preferentially synthesized in hairy roots. However, no metabolites which preferentially accumulated in the cell suspension culture have been successfully annotated. Overall, the accumulation pattern of secondary products and their potential intermediates were mostly

**Table 3** The accumulation of annotated metabolites involved in secondary metabolite synthesis in *O. pumila*

| Class | Annotation | Exact mass (*m/z*) | Fold change (HR/callus) | *t*-test (*P*-value) |
|---|---|---|---|---|
| Camptothecin derivatives | Strictosidine (**1**) | 530.2262 | 6.1 | 3.08E-05 |
| | Strictosamide (**2**) | 498.1994 | 98.7 | 1.27E-05 |
| | | 498.1995 | 106.0 | 6.66E-07 |
| | | 498.1996 | 303.3 | 5.96E-06 |
| | | 498.1997 | 18.0 | 1.03E-03 |
| | Pumiloside (**3**)/Lyalosidic acid (**8**) | 512.1790 | 27.8 | 3.28E-10 |
| | | 512.1792 | 35.3 | 2.42E-07 |
| | Deoxypumiloside (**4**) | 496.1841 | 24.4 | 5.66E-10 |
| | | 496.1843 | 89.6 | 1.15E-06 |
| | | 496.1844 | 35.2 | 1.66E-07 |
| | $C_{26}H_{26}N_2O_8$ (**5**) | 494.1683 | 27.6 | 6.10E-07 |
| | | 494.1686 | 10.5 | 6.91E-08 |
| | | 494.1689 | 6.4 | 1.77E-04 |
| | $C_{20}H_{16}N_2O_3$ (**6**) | 332.1154 | 5.7 | 7.88E-04 |
| | | 332.1165 | 32.1 | 5.79E-07 |
| | Camptothecin (**7**) | 348.1108 | 125.6 | 7.35E-04 |
| Anthraquinone derivatives | 1-Hydroxy-2-methylanthraquinone (**9**)/3-Hydroxy-2-methylanthraquinone (**10**) | 238.0628 | 8.6 | 3.62E-04 |
| | | 238.0629 | 7.9 | 5.21E-04 |
| | 3-Hydroxyanthraquinone-2-carbaldehyde (**11**) | 252.0424 | 13.8 | 5.62E-04 |
| | 1-Hydroxy-2-hydroxymethylanthraquinone (**12**)/3-Hydroxy-2-hydroxymethylanthraquinone (**13**)/1,3-Dihydroxy-2-methylanthraquinone (**14**) | 254.0576 | 13.3 | 8.06E-05 |
| | | 254.0577 | 41.3 | 2.48E-05 |
| | | 254.0580 | 54.5 | 5.82E-07 |
| | 2-Hydroxymethyl-3-methoxyanthraquinone (**15**) | 268.0732 | 27.5 | 3.94E-05 |
| | | 268.0736 | 14.1 | 2.93E-06 |
| | 1,3-Dihydroxy-2-hydroxymethylanthraquinone (**16**) | 270.0523 | 2.5 | 6.73E-03 |
| | | 270.0525 | 9.9 | 3.75E-04 |
| | | 270.0529 | 23.9 | 4.93E-05 |
| | 1-Hydroxy-2-hydroxymethyl-3-methoxyanthraquinone (**17**)/1,3-Dihydroxy-2-methoxymethylanthraquinone (**18**) | 284.0681 | 27.7 | 2.08E-04 |
| | | 284.0685 | 51.4 | 1.37E-07 |
| | | 284.0685 | 11.0 | 4.48E-04 |
| | 2-*n*-buthoxymethyl-1,3-dihydroxyanthraquinone (**19**) | 326.1156 | 23.8 | 1.61E-06 |
| | Lucidin 3-*O*-primeveroside (**20**) | 564.1480 | 24.2 | 1.02E-05 |
| | | 564.1483 | 12.6 | 8.40E-11 |
| Others | Chlorogenic acid (3-*O*-Caffeoylquinic acid) (**21**) | 354.0949 | 63.6 | 1.25E-03 |
| | | 354.0952 | 45.7 | 9.89E-09 |
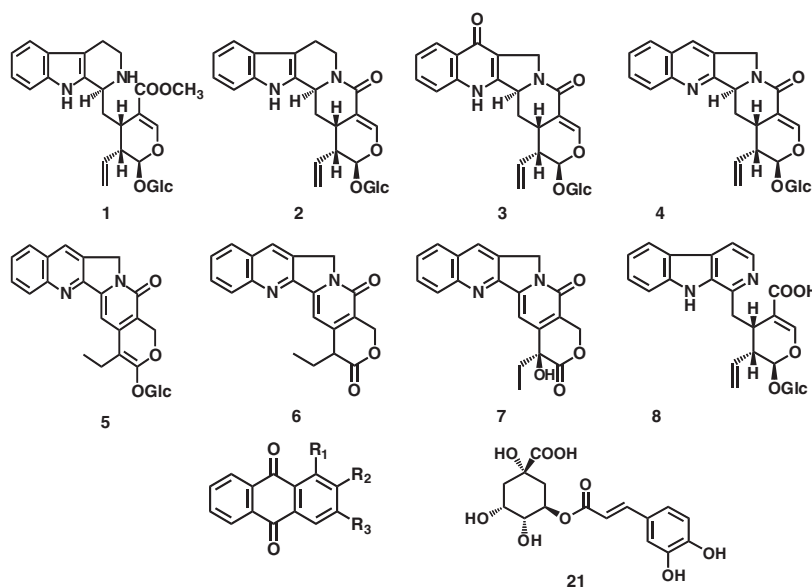


9  R₁= OH, R₂= CH₃, R₃= H
10 R₁= H, R₂= CH₃, R₃= OH
11 R₁= H, R₂= CHO, R₃= OH
12 R₁= OH, R₂= CH₂OH, R₃= H
13 R₁= H, R₂= CH₂OH, R₃= OH
14 R₁= OH, R₂= CH₃, R₃= OH
15 R₁= H, R₂= CH₂OH, R₃= OCH₃
16 R₁= OH, R₂= CH₂OH, R₃= OH
17 R₁= OH, R₂= CH₂OH, R₃= OCH₃
18 R₁= OH, R₂= CH₂OCH₃, R₃= OH
19 R₁= OH, R₂= CH₂OCH₂CH₂CH₂CH₃, R₃= OH
20 R₁= OH, R₂= CH₂OH, R₃= OGlcXyl

The annotation of metabolites was performed on the basis of observed *m/z* values from in-fusion FT-ICR-MS analyses in comparison with theoretical values. HR, hairy roots.

correlated with the expression of their biosynthetic genes in the transcriptome data.

## Discussion

Deep transcriptome analysis for medicinal plants provides an opportunity to discover genes involved in the pathways that lead to the synthesis of plant specialized products. If biological resources, e.g. mutants, cell lines or developmental tissues, are available they can facilitate pathway discovery through the differential accumulation of a compound of interest. In particular, if one can combine transcriptome data with metabolic profiling data, one can generate correlations of genes and metabolites in given biological systems on a whole-genome and metabolome scale. Such an integrated 'omics' approach can greatly facilitate prediction of new and novel genes committed in the pathways of interest (Saito et al. 2008, Fukushima et al. 2009, Saito and Matsuda 2010). In this study, data sets for predicting the genes and metabolites involved in the camptothecin and anthraquinone biosynthetic pathways in *O. pumila* hairy roots were generated. These data sets will be useful for further discovery of the genes and metabolites committed in these secondary metabolic pathways as discussed in other articles of this special issue (see Muranaka and Saito; Ramilowski et al.). In particular, generation of a hypothesis for the candidate genes involved in the particular biosynthetic pathways can be achieved by a comparison of gene expression and metabolite accumulation, as demonstrated recently by a similar approach (Geu-Flores et al. 2012; Yonekura-Sakakibara et al. 2013; Higashi and Saito 2013).

Regarding camptothecin biosynthesis, the steps after strictosamide to camptothecin are not well characterized in terms of which chemical reactions, intermediary metabolites and enzymes are involved. Some candidates have been proposed by recent gene knockdown experiments (Asano et al. 2012). The present study on untargeted metabolic profiling by FT-ICR-MS supports the presence of possible intermediates (**Fig. 1**). The post-strictosamide events of camptothecin biosynthesis may include a number of steps such as oxidative ring expansion, reduction, dehydration, isomerization (double bond migration), deglycosylation and hydroxylation. In addition to these structural conversion reactions, the mechanisms of transport across organelles and even cells may be required. The unigenes which encode these proteins potentially involved in these steps could be preferentially expressed in the hairy roots.

Cyt P450s are presumably involved in some of the steps post-strictosamide, and 229 unigenes exhibited similarity to Cyt P450s (**Supplementary Table S4**). Of these, 130 unigenes were classified into 33 P450 subfamilies and were up-regulated in hairy roots compared with cell suspension culture. The CYP71B subfamily was the most abundant subfamily (21 unigenes), followed by the CYP6C subfamily (13 unigenes), the CYP716A subfamily (12 unigenes) and the CYP71A subfamily (11 unigenes). As P450s belonging to these subfamilies are

known to catalyze certain reactions of plant secondary metabolism (Mizutani and Ohta 2010, Bak et al. 2011), these contigs are candidates to encode the Cyt P450s that catalyze the reactions involved in the post-strictosamide events of camptothecin biosynthesis. Similarly, unigenes encoding glycosidase-like proteins were examined. A total of 1,245 unigenes exhibited similarity with glycosidase-like proteins, and 93 unigenes were up-regulated in hairy roots compared with cell suspension culture. These unigenes belong to 21 subfamilies classified by the CAZy database (http://www.cazy.org/Glycoside-Hydrolases.html) (**Supplementary Table S4**). The GH1 family was the most abundant family (38 unigenes), followed by GH3 (15 unigenes). There is a gene (Contig8678) with sequence similarity to the reported strictosidine and raucaffricine acting as substrates for β-glucosidases from *Rauvolfia serpentina* (Xia et al. 2011). This gene might be involved in the deglycosylation reaction in the biosynthesis of camptothecin or degradation of strictosidine.

With respect to the biosynthesis of anthraquinones in *O. pumila*, little is known regarding the genes in this pathway, though the biosynthetic pathway has been well defined by tracer experiments in cell culture (Leistner 1985, Han et al. 2001). Our present study has the potential to narrow down the genes involved in this pathway by a combination of deep transcriptome and metabolome analysis. Functional characterization of these delineated genes, together with the detailed investigation of anthraquinone biosynthesis, will be undertaken in a future study.

Several national or international genomics and functional genomics projects focused on medicinal plants have emerged recently, including National Institutes of Health (http://medicinalplantgenomics.msu.edu/), Canada Genome (http://www.phytometasyn.com/) and 1KP Plant (http://www.onekp.com/project.html) initiatives. Accessing these data sets will be useful for further fine delineation of candidate genes, such as comparison of gene families in *C. acuminata* (Sun et al. 2011), *Cathranthus roseus* and *R. serpentina*, all of which produce monoterpenoid indole alkaloids.

## Materials and Methods

### Plant materials

Hairy roots of *O. pumila* were induced from aseptic plants as described previously (Saito et al. 2001). The callus and cell suspension cultures of *O. pumila* were induced from the hairy roots by addition of phytohormones as described elsewhere (Asano et al. 2012). The cell suspension cultures and the hairy roots were maintained in Gamborg B5 liquid medium (Gamborg et al. 1968) containing 2% sucrose with $0.5\,\mu M$ naphthaleneacetic acid and $5\,\mu M$ $N^6$-benzyladenine for cell suspension cultures or without these phytohormones for hairy roots at $25°C$ on a rotary shaker (125 r.p.m. for cell suspension cultures or 80 r.p.m. for hairy roots) under dark conditions.

## Construction of a full-length cDNA library and sequence data trimming

A full-length cDNA library was constructed from the poly(A)$^+$ RNA obtained from the hairy roots of *O. pumila* (3 weeks old) as described previously (Carninci et al. 2001) and sequenced on ABI3730xl sequencers (Life Technologies) (Taji et al. 2008). Bases and quality values were determined using the Phred program (Ewing and Green 1998, Ewing et al. 1998), and low quality regions (Phred quality score <20, and >20 bases repeated) were discarded. Vector sequences were identified using the cross_match program (Ewing and Green 1998, Ewing et al. 1998) with the parameters '-minmatch 10 -minscore 20 parameters'. Sequences shorter than 100 bases were removed. ESTs are available in the DNA Data Bank of Japan (DDBJ) (Kodama et al. 2012) (accession Nos. HX703880 to HX739052).

## Deep transcriptome analysis with the Illumina platform

Poly(A)$^+$ RNA was isolated from 3-week-old hairy roots and cell suspension cultures of *O. pumila*. cDNA libraries were constructed using the Illumina TruSeq™ Sample Preparation Kits and sequenced using the paired-end method with an Illumina HiSeq 2000/cBOT platform. Each fragment was sequenced to a read length of 90 nucleotides from each end; 29,682,050 hairy root-derived and 24,617,708 cell suspension culture-derived reads were generated. The Illumina reads have been deposited in the DDBJ Sequence Read Archive as the accession No. DRA000930.

## Assembly of deep transcriptome data and generation of a hybrid transcriptome unigene set

A schematic workflow of the computational analysis to assemble the *O. pumila* transcriptome data is shown in **Supplementary Fig. S1**. The FASTX tool kit was used to pre-process paired-end reads of 90 bp length. Based on the quality score distributions, reads were trimmed to 75 bp and then homopolymers were removed and only sequences with a quality score ≥20 and a minimum length of 70 bp were retained (Blankenberg et al. 2010); Version 0.0.13, http://hannonlab.cshl.edu/fastx_toolkit). A de novo transcriptome assembly of cleaned paired-end reads was performed using the Oases assembler with a k-mer length of 31 (Schulz et al. 2012). Oases-derived contigs and the ESTs from the full-length cDNAs derived from Sanger sequencing were assembled into a non-redundant transcript data set (unigene) using the CAP3 program (Huang and Madan 1999) with a parameter of sequence identity >90% (−p 90). Transcripts in the hybrid assembly will be available to download via our site, http://ngs-data-archive.psc.riken.jp/pub/ophiorrhiza_pumila/download/.

## Differential expression analysis of RNA seq data

The Illumina reads were mapped to the unigene set using the Bowtie program with a default setting in the single end mode. Transcripts were quantified and differential expression assessed using Cufflinks v.0.9.0 (Trapnell et al. 2010). The cuffmerge command was used to calculate FPKM values of each gene in the tissues of cell suspension culture or hairy roots. The cuffdiff command was used to test and identify differentially expressed unigenes with an FDR-corrected *P*-value of <0.05. Additionally, unigenes specifically expressed in either tissue with significant expression level were annotated as differentially expressed genes.

## Functional annotations

To predict the function of the unigenes, we performed a similarity search of the unique gene set as query data using the BLASTx program (E-value <10$^{-5}$) (Altschul et al. 1997) against protein databases of the NCBI nr and the *A. thaliana* proteome (TAIR10) (Lamesch et al. 2012). The BLAST result with NCBI nr (Benson et al. 2012) proteins was applied to identify putative Cyt P450-encoding genes in *O. pumila* based on sequence similarity. The BLAST result with *A. thaliana* proteins was applied to identify enriched GO terms using the GOrilla web server (Eden et al. 2009) as well as to identify putative glycosidase-encoding genes by reference to *A. thaliana* glycosidase genes classified in the CAZy database (Cantarel et al. 2009).

## Untargeted metabolomics analysis

Untargeted metabolomic analyses were carried out by in-fusion FT-ICR-MS method. Briefly, high, middle and non-polar extracts of plant tissues (hairy roots and cell suspension cultures; 3 weeks old) were subjected to triplicate analyses with FT-ICR-MS (APEX III, Bruker Daltonics) as described previously (Aharoni et al. 2002, Hirai et al. 2004, Tohge et al. 2005). Data were processed using DISCOVArray (Phenomenome Discoveries). Metabolite peaks were annotated based on the elemental composition calculated from accurate *m/z* values followed by a metabolite search with databases (Hocking 1997, Afendi et al. 2012). The fold changes of the peaks were calculated based on the average signal intensities obtained from hairy roots and cell suspension cultures, respectively.

# References

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K. et al. (2012) KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant Cell Physiol.* 53: e1.

Aharoni, A., Ric de Vos, C.H., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., Bino, R. et al. (2002) Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. *OMICS* 6: 217–234.

Aimi, N., Hoshino, H., Nishimura, M., Sakai, S.-i. and Haginiwa, J. (1990) Chaboside, first natural glycocamptothecin found from *Ophiorrhiza pumila*. *Tetrahedron Lett.* 31: 5169–5172.

Aimi, N., Nishimura, M., Miwa, A., Hoshino, H., Sakai, S.-i. and Haginiwa, J. (1989) Pumiloside and deoxypumiloside; plausible intermediates of camptothecin biosynthesis. *Tetrahedron Lett* 30: 4991–4994.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.

Arbain, D., Putra, D. and Sargent, M. (1993) The alkaloids of *Ophiorrhiza filistipula*. *Aust. J. Chem.* 46: 977–985.

Asano, T., Kobayashi, K., Kashihara, E., Sudo, H., Sasaki, R., Iijima, Y. et al. (2012) Suppression of camptothecin biosynthetic genes results in metabolic modification of secondary products in hairy roots of Ophiorrhiza pumila. *Phytochemistry* (in press) doi.org/10.1016/j.phytochem.2012.04.019.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S. et al. (2011) Cytochromes P450. *The Arabidopsis Book*: e0144.

Benson, D.A., Karsch-Mizrachi, I., Clark, K., Lipman, D.J., Ostell, J. and Sayers, E.W. (2012) GenBank. *Nucleic Acids Res.* 40: D48–D53.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J. and Nekrutenko, A. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26: 1783–1785.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: D233–D238.

Carninci, P., Shibata, Y., Hayatsu, N., Itoh, M., Shiraki, T., Hirozane, T. et al. (2001) Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis. *Genomics* 77: 79–90.

Chae, L., Lee, I., Shin, J. and Rhee, S.Y. (2012) Towards understanding how molecular networks evolve in plants. *Curr. Opin. Plant Biol.* 15: 177–184.

Chan, H.-H., Li, C.-Y., Damu, A.G. and Wu, T.-S. (2005) Anthraquinones from *Ophiorrhiza hayatana* OHWI. *Chem. Pharm. Bulll* 53: 1232–1235.

Darwin, S.P. (1976) The Pacific species of *Ophiorrhiza* L. (Rubiaceae). *Lyonia* 1: 47–102.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48.

Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8: 186–194.

Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175–185.

Fukushima, A., Kusano, M., Redestig, H., Arita, M. and Saito, K. (2009) Integrated omics approaches in plant systems biology. *Curr. Opin. Chem. Biol.* 13: 532–538.

Gamborg, O.L., Miller, R.A. and Ojima, K. (1968) Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.* 50: 151–158.

Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C. et al. (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* 492: 138–142.

Gross, J., Cho, W.K., Lezhneva, L., Falk, J., Krupinska, K., Shinozaki, K. et al. (2006) A plant locus essential for phylloquinone (vitamin K1) biosynthesis originated from a fusion of four eubacterial genes. *J. Biol. Chem.* 281: 17189–17196.

Han, Y.-S., Heijden, R.v.d., Lefeber, A.W.M., Erkelens, C. and Verpoorte, R. (2002) Biosynthesis of anthraquinones in cell cultures of Cinchona 'Robusta' proceeds via the methylerythritol 4-phosphate pathway. *Phytochemistry* 59: 45–55.

Han, Y.-S., Van der Heijden, R. and Verpoorte, R. (2001) Biosynthesis of anthraquinones in cell cultures of the Rubiaceae. *Plant Cell Tissue Org. Cult.* 67: 201–220.

Higashi, Y. and Saito, K. (2013) Network analysis for gene discovery in plant-specialized metabolism. *Plant, Cell Environ.* in press, doi: 10.1111/pce.12069.

Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y. et al. (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *J. Biol. Chem.* 280: 25590–25595.

Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A. et al. (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. USA* 104: 6478–6483.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M. et al. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* 101: 10205–10210.

Hocking, G.M (1997) *In* A Dictionary of Natural Products: Terms in the Field of Pharmacognosy Relating to Natural Medicinal and Pharmaceutical Materials and the Plants, Animals, and Minerals from Which They are Derived. Medford, NJ, Plexus Publishing.

Hsiang, Y.H., Hertzberg, R., Hecht, S. and Liu, L.F. (1985) Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* 260: 14873–14878.

Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* 9: 868–877.

Hutchinson, C.R., Heckendorf, A.H., Straughn, J.L., Daddona, P.E. and Cane, D.E. (1979) Biosynthesis of camptothecin. 3. Definition of strictosamide as the penultimate biosynthetic precursor assisted by carbon-13 and deuterium NMR spectroscopy. *J. Amer. Chem. Soc.* 101: 3358–3369.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38: D355–D360.

Kitajima, M., Fischer, U., Nakamura, M., Ohsawa, M., Ueno, M., Takayama, H. et al. (1998) Anthraquinones from *Ophiorrhiza pumila* tissue and cell cultures. *Phytochemistry* 48: 107–111.

Kitajima, M., Fujii, N., Yoshino, F., Sudo, H., Saito, K., Aimi, N. et al. (2005) Camptothecins and two new monoterpene glucosides from *Ophiorrhiza liukiuensis*. *Chem. Pharm. Bull.* 53: 1355–1358.

Kodama, Y., Mashima, J., Kaminuma, E., Gojobori, T., Ogasawara, O., Takagi, T. et al. (2012) The DNA Data Bank of Japan launches a new resource, the DDBJ Omics Archive of functional genomics experiments. *Nucleic Acids Res.* 40: D38–D42.

Koshiro, Y., Jackson, M.C., Katahira, R., Wang, M.L., Nagai, C. and Ashihara, H. (2007) Biosynthesis of chlorogenic acids in growing and ripening fruits of *Coffea arabica* and *Coffea canephora* plants. *Z. Naturforsch. C* 62: 731–742.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C. et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202–D1210.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10: R25.

Leistner, E. (1985) Biosynthesis of chorismate-derived quinones in plant cell cultures. *In* Primary and Secondary Metabolism of Plant Cell Cultures. Edited by Neumann, K.H., Barz, W. and Reinhardt, E. pp. 215–224. Springer, Berlin.

Matsuda, F., Hirai, M.Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N.J. et al. (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant Physiol.* 152: 566–578.

Mizutani, M. and Ohta, D. (2010) Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol* 61: 291–315.

Okazaki, Y., Shimojima, M., Sawada, Y., Toyooka, K., Narisawa, T., Mochida, K. et al. (2009) A chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. *Plant Cell* 21: 892–909.

Ozsolak, F. and Milos, P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12: 87–98.

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics - Majority report by precogs. *Trends Plant Sci.* 13: 36–43.

Saito, K. and Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61: 463–489.

Saito, K.S., Sudo, H.S., Yamazaki, M.Y., Koseki-Nakamura, M.K.-N., Kitajima, M.K., Takayama, H.T. et al. (2001) Feasible production of camptothecin by hairy root culture of *Ophiorrhiza pumila*. *Plant Cell Rep.* 20: 267–271.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28: 1086–1092.

Sirikantaramas, S., Asano, T., Sudo, H., Yamazaki, M. and Saito, K. (2007) Camptothecin: therapeutic potential and biotechnology. *Curr. Pharm. Biotechnol.* 8: 196–202.

Sirikantaramas, S., Yamazaki, M. and Saito, K. (2008) Mutations in topoisomerase I as a self-resistance mechanism coevolved with the production of the anticancer alkaloid camptothecin in plants. *Proc. Natl Acad. Sci. USA* 105: 6782–6786.

Sun, Y., Luo, H., Li, Y., Sun, C., Song, J., Niu, Y. et al. (2011) Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics* 12: 533.

Taji, T., Sakurai, T., Mochida, K., Ishiwata, A., Kurotani, A., Totoki, Y. et al. (2008) Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, Thellungiella halophila. *BMC Plant Biol.* 8: 115.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M. et al. (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J.* 42: 218–235.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J. et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28: 511–515.

Viraporn, V., Yamazaki, M., Saito, K., Denduangboripant, J., Chayamarit, K., Chuanasa, T. et al. (2011) Correlation of camptothecin-producing ability and phylogenetic relationship in the genus Ophiorrhiza. *Planta Med.* 77: 759–764.

Wall, M.E., Wani, M.C., Cook, C.E., Palmer, K.H., McPhail, A.T. and Sim, G.A. (1966) Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from *Camptotheca acuminata*. *J. Amer. Chem. Soc* 88: 3888–3890.

Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57–63.

Xia, L., Ruppert, M., Wang, M., Panjikar, S., Lin, H., Rajendran, C. et al. (2011) Structures of alkaloid biosynthetic glucosidases decode substrate specificity. *ACS Chem. Biol.* 7: 226–234.

Yamazaki, Y., Kitajima, M., Arita, M., Takayama, H., Sudo, H., Yamazaki, M. et al. (2004) Biosynthesis of camptothecin. In silico and in vivo tracer study from [1-13C]glucose. *Plant Physiol.* 134: 161–170.

Yamazaki, Y., Sudo, H., Yamazaki, M., Aimi, N. and Saito, K. (2003a) Camptothecin biosynthetic genes in hairy roots of *Ophiorrhiza pumila*: cloning, characterization and differential expression in tissues and by stress compounds. *Plant Cell Physiol.* 44: 395–403.

Yamazaki, Y., Urano, A., Sudo, H., Kitajima, M., Takayama, H., Yamazaki, M. et al. (2003b) Metabolite profiling of alkaloids and strictosidine synthase activity in camptothecin producing plants. *Phytochemistry* 62: 461–470.

Yonekura-Sakakibara, K., Fukushima, A., Nakabayashi, R., Hanada, K., Matsuda, F., Sugawara, S. et al. (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* 69: 154–167.

Yonekura-Sakakibara, K., Fukushima, A. and Saito, K. (2013) Transcriptome data modeling for targeted plant metabolic engineering. *Curr. Opin. Biotech.* 24: 285–290.

Yonekura-Sakakibara, K. and Saito, K. (2009) Functional genomics for plant natural product biosynthesis. *Nat. Prod. Rep.* 26: 1466–1487.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R. et al. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* 20: 2160–2176.

Yonekura-Sakakibara, K., Tohge, T., Niida, R. and Saito, K. (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* 282: 14932–14941.