Journal of Microbiology (2017) Vol. 55, No. 11, pp. 871–876
DOI 10.1007/s12275-017-7173-1
eISSN 1976-3794
pISSN 1225-8873

De novo transcriptome assembly and characterization of the 10-hydroxycamptothecin-producing *Xylaria* sp. M71 following salicylic acid treatment[§]

Xiaowei Ding[1,2], Kaihui Liu[2]*, Yonggui Zhang[2], and Feihu Liu[1]*

[1]*School of Agriculture, Yunnan University, Kunming 650091, P. R. China*
[2]*School of Biological Science and Engineering, Shaanxi University of Technology, Hanzhong 723001, P. R. China*

(Received Apr 21, 2017 / Revised Sep 5, 2017 / Accepted Oct 9, 2017)

In the present study, we identified genes that are putatively involved in the production of fungal 10-hydroxycamptothecin via transcriptome sequencing and characterization of the *Xylaria* sp. M71 treated with salicylic acid (SA). A total of 60,664,200 raw reads were assembled into 26,044 unigenes. BLAST assigned 8,767 (33.7%) and 10,840 (41.6%) unigenes to 40 Gene Ontology (GO) annotations and 108 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, respectively. A total of 3,713 unigenes comprising 1,504 upregulated and 2,209 downregulated unigenes were found to be differentially expressed between SA-induced and control fungi. Based on the camptothecin biosynthesis pathway in plants, 13 functional genes of *Xylaria* sp. M71 were mapped to the mevalonate (MVA) pathway, suggesting that the fungal 10-hydroxycamptothecin is produced via the MVA pathway. In summary, analysis of the *Xylaria* sp. M71 transcriptome allowed the identification of unigenes that are putatively involved in 10-hydroxycamptothecin biosynthesis in fungi.

Keywords: transcriptomic analysis, *Xylaria* sp. M71, 10-hydroxycamptothecin

Introduction

Endophytic fungi are considered a rich source of phytochemical drugs and plant-derived anticancer compounds, such as camptothecin (CPT) (Kusari *et al*., 2009). CPT and its derivatives were originally discovered in *Camptotheca acuminate*. These anticancer terpenoids were found to be produced by fungal endophytes associated with *C. acuminata*, such as *Fusarium solani* (Kusari *et al*., 2009) and *Xylaria* sp. (Liu *et al*., 2010). However, dramatic reductions in the yield of fungus-derived CPT analogs after repeated endophyte subculturing represents a great challenge for establishing and maintaining a biosynthesis system for CPT analogs (Kusari *et al*., 2009). Recently, elicitor-induced reprogramming of metabolic pathways has been demonstrated to be a promising strategy for increasing the yield of microbial pharmaceuticals (Robin, 2011).

Accumulating evidence has indicated that elicitor compounds can markedly enhance the production of secondary metabolites by eukaryotes. For example, 1,3-diaminopropane was demonstrated to increase penicillin production by *Penicillium chrysogenum* (Martín *et al*., 2011). Salicylic acid (SA) has been shown to promote the accumulation of benzophenanthridine alkaloids in plant suspension cultures via regulation of a specific signal transduction pathway (Cho *et al*., 2008). In addition, SA was found to significantly increase the 10-hydroxycamptothecin (10-HCPT) production by the endophytic fungus *Xylaria* sp. M71 (Liu *et al*., 2010). However, the SA-regulated genes and metabolic pathways that are involved in fungal 10-HCPT production remain unknown.

Transcriptome sequencing is a powerful technique for analyzing functional genes and metabolic pathways in microorganisms (Hoang *et al*., 2016). In the present study, transcriptional profiles of *Xylaria* sp. M71 both in the presence and absence of SA were obtained via Illumina RNA sequencing. A total of 3,713 differential expressed unigenes (DEGs) were identified in SA-induced *Xylaria* sp. M71, out of which 13 DEGs were putatively involved in 10-HCPT biosynthesis. Our findings provide a basis for enhancing 10-HCPT production in fungi through genetic engineering.

Materials and Methods

Organism and culture conditions

Xylaria sp. M71, which produces 10-hydroxycamptothecin, was provided by the Shaanxi Provincial Engineering Research Center of Edible and Medicinal Microbes. M71 was inoculated onto 300 ml flasks containing Sabouraud dextrose liquid media (40 g of glucose, 10 g of peptone, pH 6.0) at 28°C for 7 days. Some fungal cultures were added with 0.1 mmol salicylic acid (SA) at 28°C for 2 h and were marked as Xylaria1A. Control samples, marked as Xylaria0A, were maintained under the same experimental conditions but without the addition of SA. The mycelia of *Xylaria* sp. M71 were then harvested for total RNA extraction.

*For correspondence. (K. Liu) E-mail: kaihhui168@163.com; Tel. & Fax: +86-916-2641661 / (F. Liu) E-mail: dmzpynu@126.com; Tel. & Fax: +86-871-65031539
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.
Copyright ⓒ 2017, The Microbiological Society of Korea

Table 1. Primers used in this study

Genes	Primers (5′-3′)
AACT	5′-TACTCTGTTCTCGAGAGGA-3′ 5′-TCGTCAACTCATTCTTGAT-3′
GGPS	5′-TCCTCGACAGGCTACATTA-3′ 5′-AGGTATAGCAGTAGTCGAG-3′

RNA extraction, library construction, and transcriptome sequencing

Total RNA was extracted using the Trizol reagent (Invitrogen). Extracted RNA were quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies). Triplicate extractions were performed for each sample, and the obtained extracts were pooled for library construction. cDNA libraries were constructed using the purified mRNA fragments according to the manufacturer's instructions (Illumina, Inc.). Sequencing was performed on the Illumina HiSeq 2000 platform at the Beijing Genomics Institute. Raw reads were deposited in the Short Read Archive database at the National Center for Biotechnology Information (NCBI) under the accession numbers SRS1463634 and SRS1463635.

De novo assembly by Trinity

Raw reads were filtered by discarding the adapter sequences and low-quality sequences. Clean reads were used as input for *de novo* assembly using Trinity (Haas *et al.*, 2013). Briefly, reads with a certain overlap length were combined to produce contigs. Trinity connects the contigs into sequences that could not be extended on either end and are considered as transcripts. Homologous clustering of transcripts was performed and served as the basis for defining unigenes. Unigenes were then grouped into clusters and singletons via gene family clustering.

Functional annotation and differential expression analysis

Unigenes were aligned against the NCBI non-redundant database (NR), Swiss-Prot protein database (Swiss-Prot), Kyoto Encyclopedia of Gene and Genomes database (KEGG) (Moriya *et al.*, 2007), and the Clusters of Orthologous Groups of proteins database (COG) using BLASTx (*E*-value < 10^{-5}). Gene Ontology (GO) analysis was performed using the Blast2GO software (Conesa *et al.*, 2005). GO functional classification was performed using WEGO software (Ye *et al.*, 2006). Differentially expressed unigenes in SA-stressed vs. uninduced (control) samples were filtered using a threshold of false discovery rate (FDR) ≤ 0.001 and an absolute log2 ratio ≥ 1. DEGs were further mapped onto known pathways through KEGG pathway annotation.

Quantitative real-time PCR (qPCR) analysis

qPCR analysis was performed on a 7500 Real-Time PCR System (Applied Biosystems) to analyze the relative expression levels of two DEGs that are putatively involved in 10-HCPT biosynthesis. PCR reactions were carried out in triplicate for each gene using endogenous 18S rRNA as the reference gene. Gene-specific primers are provided in Table 1.

Results and Discussion

Illumina sequencing and *de novo* transcriptome assembly

Illumina sequencing generated a total of 60,664,200 and 70,183,240 raw reads from the SA-induced and control cultures, respectively. After removal of low-quality bases, adaptors, and ambiguous reads, 52,776,108 and 53,115,684 clean reads were obtained, respectively, from the SA-induced and control *Xylaria* sp. M71 samples. The reads had a Q20 value greater than 97% (percentage of sequences with fragment lengths ≥ 20 nt) (Table 2). All clean sequences were assembled into 38,626 contigs with a mean length of 512 nucleotides (nt) for the SA-induced samples and 42,305 contigs with a mean length of 428 nt for the control samples. The majority of contigs from the SA-treated samples and controls (64.33 and 68.47%, respectively) had lengths ranging from 200 to 400 nt. Contigs were assembled into 31,897 unigenes for the SA-induced *Xylaria* sp. M71 samples and 34,561 unigenes for the control *Xylaria* sp. M71 samples, with mean lengths of 653 and 520 nt and N50 lengths of 1,366 and 1,061 nt, respectively (Table 3). Further clustering and assembly returned a total of 26,044 non-redundant unigenes. The analysis showed significant differences in the transcriptomes of the SA-induced and the control samples, suggesting that gene expression profiles were significantly influenced by SA treatment.

Functional annotation

Approximately 68% of unigenes (17,722) were assigned by

Table 2. Throughput and quality of RNA-Seq reads from *Xylaria* sp. M71 samples

	Xylaria1A	Xylaria0A
Total reads	60,664,200	70,183,240
Clean reads	52,776,108	53,115,684
Q20 percentage[a]	97.85%	97.33%
N percentage[b]	0.00%	0.00%
GC percentage[c]	47.84%	47.51%

[a]Q20 percentage indicates the percentage of sequences with fragment lengths greater than 20 nt; b N percentage indicates proportion of unknown nucleotides in clean reads. The percentage of nucleotides without being sequenced successfully, c GC percentage indicates the percentage of G and C in total clean bases.

Table 3. Summary of the *Xylaria* sp. M71 transcriptome assembly

	Xylaria1A	Xylaria0A
Number of contigs	38,626	42,305
Total length of contigs (nt)	19,762,228	18,119,839
Mean length of contig (nt)	512	428
N50 contig length(nt)	1,354	1,021
Total unigenes[a]	31,897	34,561
Total length of unigenes (nt)	20,819,239	17,984,083
Mean length of unigene (nt)	653	520
N50 unigene length (nt)	1,366	1,061
Distinct clusters[b]	4,701	4,561
Distinct singletons[c]	27,196	30,000

[a]Total unigenes indicates the total number of assembled unigenes; [b]Distinct clusters indicates the unigenes clustered with ≥ 70% homology with the cluster and may represent the same gene or its homologs; [c]Distinct singletons indicates the number of single genes.

BLASTx to genes from the public databases using an *E*-value threshold of 10^{-5}. Among the public databases used for annotation, the majority of unigenes (17,435) were annotated to the NR database, followed by COG, KEGG, and the Swiss-Prot databases (14,303, 10,840, and 10,648, respectively). Based on NR annotations and the *E*-value distribution, 54.7% of the sequences showed very strong homology (*E*-value < $1.0E^{-45}$) to available eukaryote sequences, and 45.3% had *E*-values ranging from $1.0E^{-45}$ to $1.0E^{-5}$ (Fig. 1A). Based on NR annotations and similarity distributions, approximately 91.5% of all BLASTx hits had sequence similarities ranging from 40% to 100%, and only 8.5% of hits had similarity values ≤ 40% (Fig. 1A). Sequence homology searching against different species returned the closest match of 46.2% to *Eutypalata* (Fig. 1A).

GO and COG classification

GO terms were assigned to the assembled unigenes based on the NR annotations. A total of 8,767 unigenes were categorized under the three main GO categories, namely, biological process, cellular component, and molecular function (Fig. 1B). Under the biological process category, the majority of unigenes were assigned under metabolic process (5,017 unigenes, 57.2%), followed by cellular process (4,692 unigenes, 53.5%). In the cellular component category, the majority of unigenes were associated with cell and cell part (2,915 unigenes, 33.2%). In the molecular function category, catalytic activity (5,099 unigenes, 58.2%) and binding (4,103 unigenes, 46.8%) represented the two most abundant subcategories. A total of 14,303 unigenes representing protein domains were

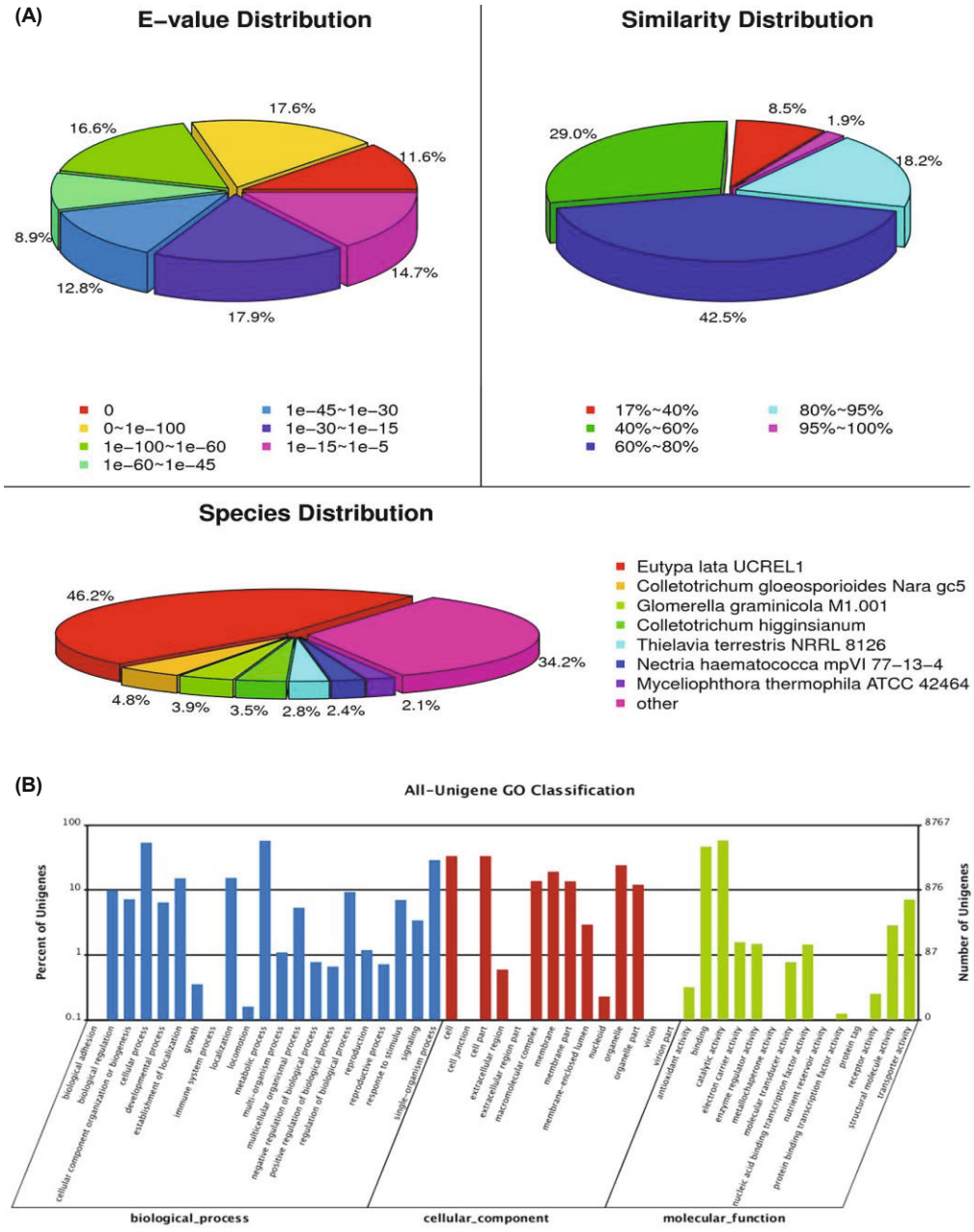

Fig. 1. Homology searching of unigenes against NR databases (A) and gene ontology classification of unigenes (B).

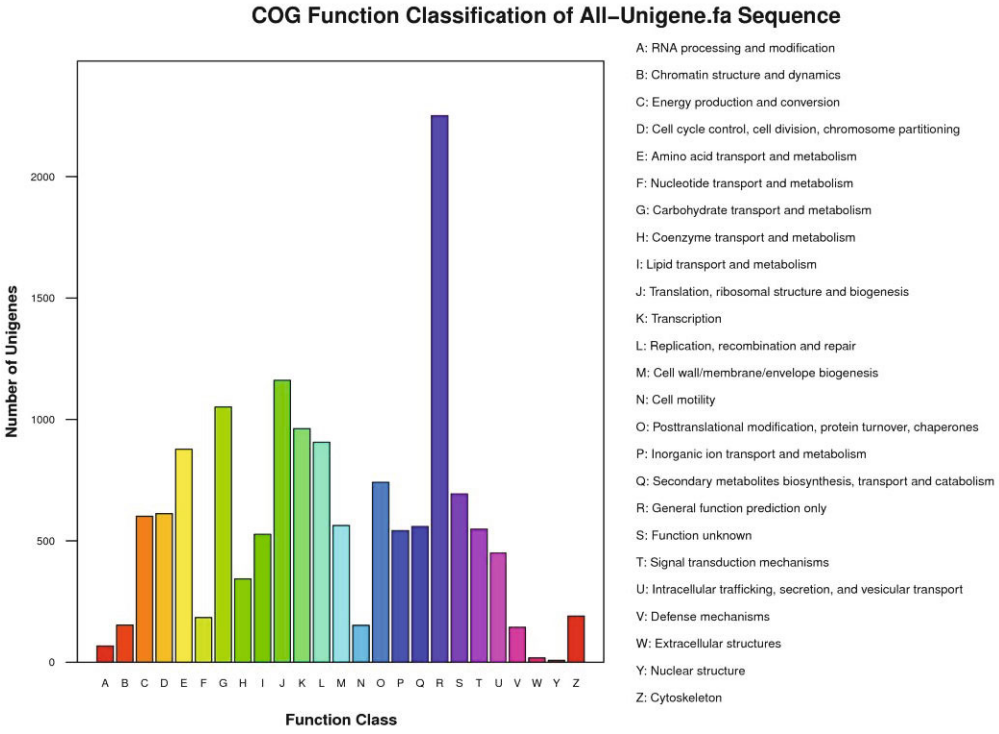

Fig. 2. COG functional classification of *Xylaria* sp. M71 transcriptomes.

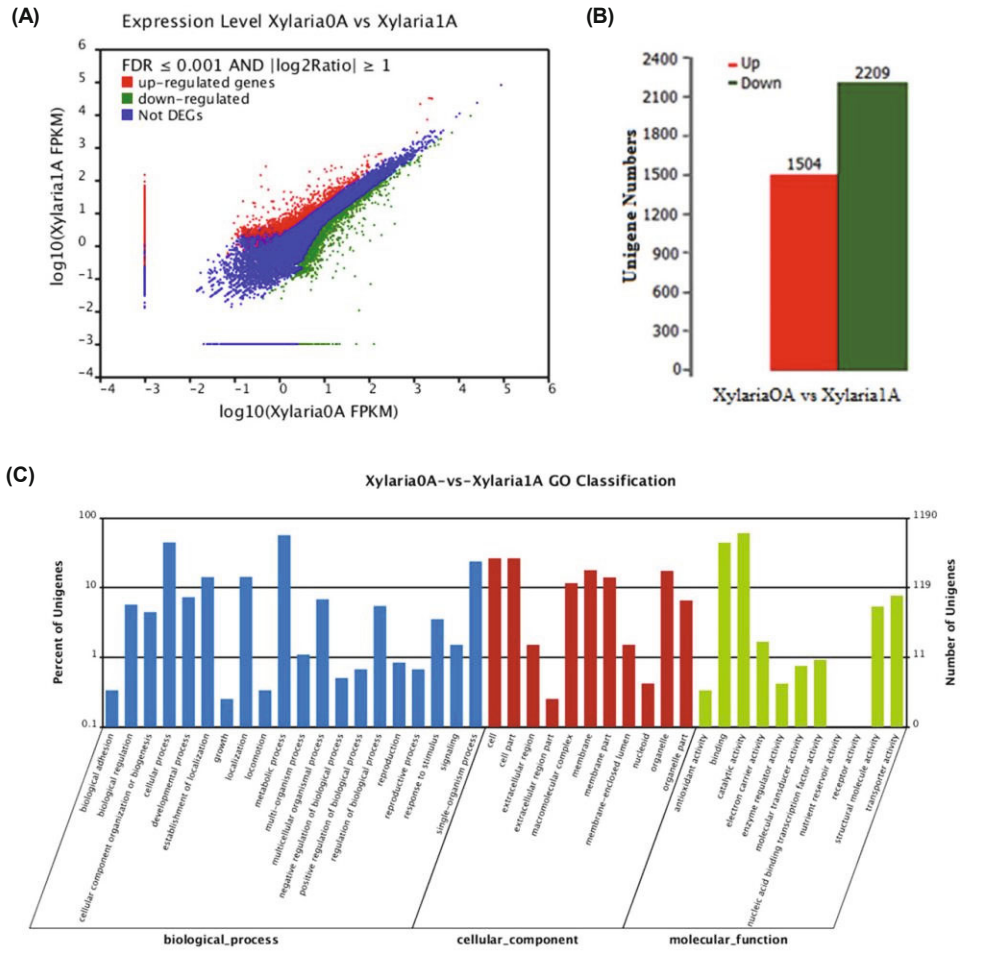

Fig. 3. Comparison of expression between Xylaria0A and Xylaria1A (A and B) and functional annotation of the differentially expressed unigenes (C).

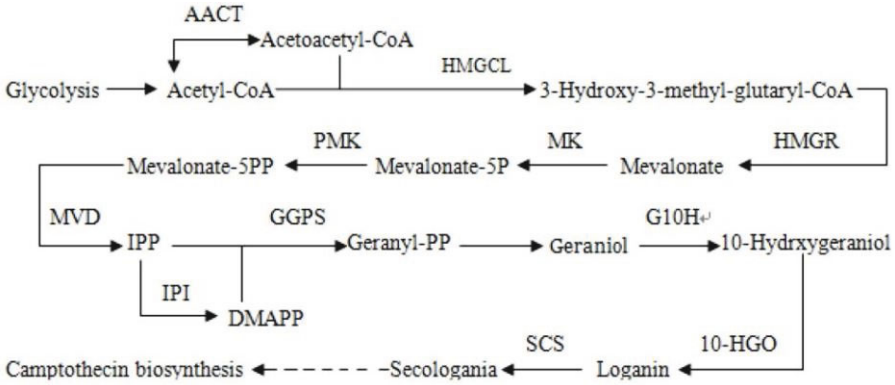

Fig. 4. **Putative 10-HCPT biosynthesis pathway in *Xylaria* sp. M71.**

categorized into 25 functional COG clusters (*E*-value ≤ $1.0E^{-5}$) (Fig. 2). The five largest COG categories were as follows: (1) general function prediction only (15.7%, 2251), (2) translation, ribosomal structure, and biogenesis (8.1%, 1161), (3) carbohydrate transport and metabolism (7.3%, 1051), (4) transcription (6.7%, 962), and (5) replication, recombination, and repair (6.3%, 906).

Differential expression and KEGG pathway analyses

In this study, significant differences in gene expression levels were considered at FDR ≤ 0.001 and log2 ratio ≥ 1 (Fig. 3A). A total of 3,713 differentially expressed genes (DEGs) comprising 1,504 upregulated and 2,209 down-regulated unigenes were identified in SA-induced *Xylaria* sp. M71 cultures when compared to control cultures, (Fig. 3B). To investigate the functions of the DEGs, all DEGs were grouped into three main categories, namely, biological process, cellular component, and molecular function. Cellular process and catalytic activity were the most abundant classification terms assigned to DEGs under the biological process category. Cell and cell part comprised the two largest subgroups under the cellular component category. Binding and catalytic activity were the dominant subgroups under the molecular function category (Fig. 3C). GO functional classification of DEGs showed that SA treatment resulted in the reprogramming of multiple cellular processes in *Xylaria* sp. M71.

To further evaluate the functions of DEGs, all DEGs were mapped to the KEGG database, out of which 1,488 DEGs were associated with 107 KEGG pathways. Consistent with the GO functional annotations, the majority of DEGs that were significantly enriched in the steroid biosynthesis pathway were downregulated, which could explain the downregulation of steroids, brassinosteroids, and vitamin D2. These compounds are known to play important roles in fungal growth, suggesting that M71 underwent a specific metabolic shift following SA treatment. Almost all the DEGs that participated in the biosynthesis of valine, leucine, and isoleucine were found to be downregulated and have been known to promote pyruvate accumulation. Correspondingly, most DEGs involved in the degradation of valine, leucine, and isoleucine were upregulated, indicating the potential targets for increasing the production of (R)-methyl-malonyl-CoA and (S)-methyl-malonyl-CoA, which are important for the biosynthesis of type II polyketide backbones and 12,14,16-membered macrolides. The above finding suggests that SA-induced changes in gene expression patterns can lead to the accumulation of polypeptides and macrolides.

Genes putatively associated with biosynthesis of fungal 10-hydroxycamptothecin

Transcriptome analyses of medical plants, such as *C. acuminata*, have provided a pathway framework of CPT biosynthesis (Sun *et al.*, 2011). However, to date, the biosynthesis pathway of CPT and its analogs in fungi remains poorly understood. Based on knowledge of the metabolic roles of CPT in plants and terpenoid indole alkaloids (TIAs) in fungi, a total of 451 unigenes represented 13 unique genes that are putatively involved in CPT biosynthesis in *Xylaria* sp. M71 (Fig. 4 and Supplementary data Table S1). These genes encode proteins, such as 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR), 10-hydroxygeraniol oxidoreductase (10-HGO), acetyl-CoA C-acetyltransferase (AACT), Geranylgeranyl pyrophosphate synthase (GGPS), 3-hydroxymethyl-3-methylglutaryl-CoA lyase (HMGCL), mevalonate kinase (MK), geraniol-10-hydroxylase (G10H), and secologanin synthase (SCS) and are known to participate in the biosynthesis of the two key precursors (secologanin and tryptamine) of CPT and its analogs in *C. acuminata*.

Among the 13 genes, four DEGs encoded G10H, 10-HGO, GGPS, and AACT. The unigenes encoding G10H and 10-HGO were downregulated. Based on qPCR analysis, one unigene that putatively encodes GGPS was upregulated, and

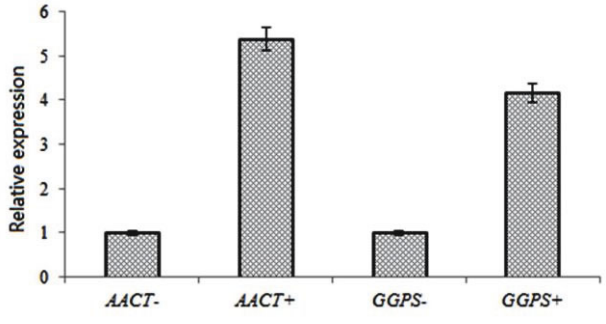

Fig. 5. **qPCR validation of *AACT* and *GGPS* gene expression levels. "–" and "+", cultures induced without and with salicylic acid, respectively.**

unigenes encoding AACT were upregulated (Fig. 5). AACT (the first rate-limiting enzyme) and GGPS are essential for the biosynthesis of terpenoid alkaloids in eukaryotes and are also indispensable for the formation of geranylgeranyl diphosphate, an important precursor of CPT (Fox *et al*., 2014). In addition, upregulation of the GGPS gene was demonstrated to increase the production of β-carotene in *Escherichia coli* (Sun *et al*., 2012). Therefore, these findings suggest that the presence of SA likely promotes AACT and GGPS production in *Xylaria* sp. M71 and is highly likely involved in the upregulation of fungal 10-HCPT as previously reported by the current authors (Liu *et al*., 2010).

The canonical pathway of the biosynthesis of terpenoid alkaloids, including CPT and its derivatives, is conserved both in plants and fungi (Sun *et al*., 2011; Yu *et al*., 2012). In this study, 13 *Xylaria* sp. M71 genes were successfully mapped to the mevalonate (MVA) pathway, which is involved in the production of strictosidine (Fig. 4 and Supplementary data Table S1), a key precursor of CPT that is indispensable for CPT production in *C. acuminata* (Sun *et al*., 2011). Besides the MVA pathway, plants can utilize the methylerythritol phosphate (MEP) and 1-deoxy-D-xylulose 5-phosphate (DXP) pathways to produce strictosidine (Burlat *et al*., 2004). However, genes associated with the MEP pathway were not identified on our analysis, thereby suggesting that 10-HCPT biosynthesis in *Xylaria* sp. M71 can utilize the mevalonate (MVA) pathway, as previously observed in terpenoid production by *Ganoderma lucidum* (Yu *et al*., 2012). G10H, SCS, and STR are considered diagnostic enzymes for strictosidine biosynthesis in TIA-producing plants; however, STR genes were not observed in the present study. Further research should employ whole genome sequencing of the *Xylaria* sp. M71 to identify other key functional genes and further elucidate the 10-HCPT biosynthesis pathway in endophytic fungi.

Conclusion

In this study, we performed transcriptome sequencing of *Xylaria* sp. M71 cultures grown in the presence and absence of SA. A total of 26,044 unigenes from SA-induced *Xylaria* sp. M71 samples and controls were functionally annotated. A total of 3,713 unigenes comprising 1,504 upregulated and 2,209 downregulated unigenes were found to be differentially expressed between the two treatment groups. A total of 13 functional genes, including *G10H*, *10-HGO*, *GGPS*, and *AACT*, were mapped to the mevalonate (MVA) pathway, which is likely to be involved in 10-HCPT production by *Xylaria* sp. M71.

Acknowledgements

We are especially grateful to the handling editor and the two anonymous reviewers for their helpful comments on the manuscript. This work was funded by the National Natural Science Foundation Program of China (No. 31100017) and the Young Science and Technology New Star Program of Shaanxi (No. 2013KJXX-76).

References

Burlat, V., Oudin, A., Courtois, M., Rideau, M., and St-Pierre, B. 2004. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J*. **1**, 131–141.

Cho, H.Y., Son, S.Y., Rhee, H.S., Yoon, S.Y., Lee-Parsons, C.W., and Park, J.M. 2008. Synergistic effects of sequential treatment with methyl jasmonate, salicylic acid and yeast extract on benzophenanthridine alkaloid accumulation and protein expression in *Eschscholtzia californica* suspension cultures. *J. Biotechnol*. **1**, 117–122.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functionalgenomics research. *Bioinformatics* **18**, 3674–3676.

Fox, A.R., Soto, G., Mozzicafreddo, M., Garcia, A.N., Cuccioloni, M., Angeletti, M., Salerno, J.C., and Ayub, N.D. 2014. Understanding the function of bacterial and eukaryotic thiolases II by integrating evolutionary and functional approaches. *Gene* **1**, 5–10.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc*. **8**, 1494–1512.

Hoang, M.H., Nguyen, C., Pham, H.Q., Nguyen, L.V., Duc, L.H., Son, L.V., Hai, T.N., Ha, C.H., Nhan, L.D., Anh, H.T.L., *et al.* 2016. Transcriptome sequencing and comparative analysis of *Schizochytrium mangrovei* PQ6 at different cultivation times. *Biotechnol. Lett*. **38**, 1781–1789.

Kusari, S., Zühlke, S., and Spiteller, M. 2009. An endophytic fungus from *Camptotheca acuminata* that produces camptothecin and analogues. *J. Nat. Prod*. **1**, 2–7.

Liu, K.H., Ding, X.W., Deng, B.W., and Chen, W.Q. 2010. 10-Hydroxycamptothecin produced by a new endophytic *Xylaria* sp., M20, from *Camptotheca acuminata*. *Biotechnol. Lett*. **5**, 689–693.

Martín, J., García-Estrada, C., Rumbero, A., Recio, E., Albillos, S.M., Ullán, R.V., and Martín, J.F. 2011. Characterization of an autoinducer of penicillin biosynthesis in *Penicillium chrysogenum*. *Appl. Environ. Microbiol*. **16**, 5688–5696.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. **35**, 182–185.

Robin, K.P. 2011. Small-molecule elicitation of microbial secondary metaboliites. *Microb. Biotechnol*. **4**, 471–478.

Sun, J., Sun, X.X., Tang, P.W., and Yuan, Q.P. 2012. Molecular cloning and functional expression of two key carotene synthetic genes derived from *Blakeslea trispora* into *E. coli* for increased β-carotene production. *Biotechnol. Lett*. **11**, 2077–2082.

Sun, Y.Z., Luo, H.M., Li, Y., Sun, C., Song, J.Y., Niu, Y.J., Zhu, Y.J., Dong, L., Lv, A. Tramontano, E., *et al.* 2011. Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics* **12**, 533.

Ye, J., Fang, L., Zheng, H.K., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., and Bolund, L. 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. **34**, W293–W297.

Yu, G.J., Wang, M., Huang, J., Yin, Y.L., Chen, Y.J., Jiang, S., Jin, Y.X., Lan, X.Q., Wong, B.H., Liang, Y., *et al.* 2012. Deep insight into the *Ganoderma lucidum* by comprehensive analysis of its transcriptome. *PLoS One* **7**, e44031.