國立交通大學

生物科技學系

博士論文

紅麴菌膽固醇抑制劑 (Monacolin K)、橘黴素及轉位子相關基因之研究

Characterization of Monacolin K, Citrinin, and Non-LTR Retrotransposon Related Genes in *Monascus pilosus*

研 究 生：陳煜沛

指導教授：曾慶平 博士

袁國芳 博士

中 華 民 國 九 十 七 年 十一 月

摘要

紅麴菌可生產各種不同的二次代謝物，包括膽固醇抑制劑(monacolin K)，橘黴素(citrinin)及色素等。紅麴菌所生產的膽固醇抑制劑 monacolin K 與 *Aspergillus terreus* 產生的 lovastatin 為結構相同的物質。橘黴素則是一種腎肝毒素，亦是針對格蘭氏陽性菌的抗菌劑，橘黴素除了紅麴菌會產生之外，於 *Aspergillus* 及 *Penicillium* 中也有發現。根據前人的研究指出，lovastatin 與 citrinin 基因群已分別在 *A. terreus* 及 *M. purpureus* 真菌中發現。因此為了解紅麴菌 monacolin K 基因群，我們從細菌人工染色體基因庫(Bacterial Artificial Chromosome library, BAC)篩選出含有 monacolin K 基因群之 BAC clones，並完成定序及註解分析，monacolin K 生合成基因群全長為 42 kb 包含 9 個基因，命名為 *mokA*-*mokI*，其與已知 lovastatin 基因群具有 54%的相似度。然而為確認所選殖基因確實參與 monacolin K 的生合成，本研究首先利用核酸酶抑制劑(aurintricarboxylic acid)建立具方便及效率性之紅麴基因轉殖技術，藉由基因轉殖篩選出 *mokA* 基因的突變株，並證實此突變株不會產生 monacolin K，進一步於紅麴菌中表現 *mokH* 基因，確認轉型株中含有兩套 *mokH* 基因，分析轉型株的 monacolin K 產量及 RNA 表現，發現其表現量皆高於野生株，因此結果明確說明所篩選的基因參與 monacolin K 的生合成。此外，選殖出 *mokA* 基因，並與 4'-phosphopantetheine transferase (PPTase) 於 *Escherichia coli* 中共同表現，發現 *E. coli* 中大量表現出色胺酸(tryptophan)的前驅物 anthranilic acid，此結果顯示 *mokA* 基因於 *E. coli* 中表現出的產物視同誘發物(elicitor)，可調控誘導 anthranilic acid 在 *E. coli* 中表現。

再者，本研究於紅麴菌的細菌人工染色體基因庫定序中發現新的 non-LTR 逆轉位子，並命名為 MRT，其序列長度約為 5.5 kb，包含有兩個開放讀架(ORFs)。此兩個開放讀架與 *gag* 及 *pol* 基因相似，並且於 *pol* 相似基因的 3'端具有很多的腺嘌呤。第一個開放讀架為 517 個氨基酸的蛋白質，包含有多量半胱氨酸的鋅指區域。第二個開放讀架為 1181 個氨基酸的蛋白質，包含有脫嘌呤/脫嘧啶核酸內

切酶、反轉錄酶、核糖核酸酶 H 及 CCHC 區域。根據 MRT non-LTR 逆轉位子的氨基酸序列所建立的親緣關係，可將其歸屬於 Tad1 群叢。南方雜交法則進一步發現只有四種紅麴菌 *M. pilosus*、*M. ruber*、*M. sanguineus* 與 *M. barkeri* 具有 MRT non-LTR 逆轉位子。除此之外，利用 β-tubulin 基因所建立的親緣關係可歸群出 MRT non-LTR 逆轉位子之存在與否。

紅麴菌廣泛應用於食品發酵及保健藥物中，因此如何分辨紅麴菌中是否含有橘黴素(citrinin)則相對性的重要，本研究分析了 18 株不同的紅麴菌種，發現只有 *M. purpureus* 與 *M. kaoliang* 具橘黴素的基因群，並且也只有此兩個菌種能偵測到橘黴素的表現量，相反的其他物種包括 *M. pilosus*、*M. ruber*、*M. barkeri*、*M. floridanus*、*M. lunisporas* 及 *M. pallens* 皆缺乏橘黴素基因群，此研究結果顯示橘黴素只有紅麴菌 *M. purpureus* 與 *M. kaoliang* 會產生，另外，利用 β-tubulin 基因所建立的親緣關係則亦可歸群出橘黴素基因群之存在與否。

Abstract

Monascus species can produce various secondary metabolites with polyketide structures as monacolin K, citrinin, and pigments. Monacolin K which inhibits cholesterol synthesis is also known as lovastatin isolated from *Aspergillus terreus*. Citrinin, the hepato-nephrotoxic agent and antibiotic against gram-positive bacteria, has been identified in *Aspergillus* and *Penicillium* spp. In previous studies, lovastatin and citrinin biosynthetic gene clusters have been characterized in *Aspergillus terreus* and *Monascus purpureus* respectively. To explore the monacolin K biosynthetic gene cluster in *M. pilosus* BCRC38072 producing monacolin K, construction of a BAC library was carried out. The putative monacolin K biosynthetic gene cluster was found within a 42-kb region in the mps01 clone. The deduced amino acid sequences encoded by the nine genes designated as *mokA* – *mokI*, which share over 54% similarity with the lovastatin biosynthetic gene cluster in *A. terreus*, were assumed to be involved in monacolin K biosynthesis. In order to characterize the putative monacolin K biosynthetic gene cluste, the valid and convenient gene transformation in *Monascus pilosus* BCRC38072 was established using nuclease inhibitor, aurintricarboxylic acid (ATA). A gene disruption constructed to replace the central part of *mokA*, a polyketide synthase gene, in wild type *M. pilosus* BCRC38072 with a hygromycin B resistance gene through homologous recombination resulted in a *mokA*-disrupted strain. The disruptant did not produce monacolin K, indicating that *mokA* encoded the PKS responsible for monacolin K biosynthesis in *M. pilosus* BCRC38072. Moreover, the transformant containing two copies of the *mokH* gene-encoded transcription factor showed higher production of monacolin K than wild type strain. Real-time RT-PCR analysis also demonstrated that the transcripts of monacolin K biosynthetic genes in the

transformant were higher than those in wild type strain. These results suggested that *mokA* and *mokH* involved in the monacolin K biosynthesis. In addition, the *mokA* gene and *sfp* gene were coexpressed in *Escherichia coli*. The *sfp* gene obtained from *Bacillus subtilis* is a phosphopantetheinyl transferase required to convert the expressed apo-PKS to its holo form. Interestingly, anthranilic acid which is the precursor of tryptophan was found. This result revealed that that expression product of polyketide synthase (*mokA*) in *E. coli* was responsible for the elicitor to regulate anthranilate flux through the anthranilate synthase gene specific to the tryptophan biosynthetic pathway.

During the BAC (mps01) sequencing of *M. pilosus* BCRC38072, a new non-LTR retrotransposon, named MRT, was discovered. The entire nucleotide sequence of the MRT element was 5.5-kb long, including two open reading frames. These two ORFs showed homologies to *gag*-like and *pol*-like gene products, and an A-rich sequence at the 3' end of *pol*-like gene. ORF1 encoded a protein of 517 amino acids and contained a cysteine-rich zinc finger motif. ORF2 encoded a protein of 1181 amino acids and contained apurinic/apyrimidinic endonuclease (APE), reverse transcriptase (RT), RNaseH domains, and a CCHC motif. The phylogenetic analyses demonstrated that the MRT element was classified into the Tad1 clade. The results of Southern hybridizations showed that MRT elements were distributed within *M. pilosus*, *M. ruber*, *M. sanguineus*, and *M. barkeri*. Also, the species of *Monascus* can be grouped by the presence or absence of MRT elements in the hybridization pattern according to phylogenetic subgroups established with the partial β-tubulin gene.

Monascus species has been widely used in food fermentation, and has shown a highly promising application in medicine development. Therefore, it is important to identify the non-citrinin producing *Monascus* strains. Eighteen *Monascus* strains

were investigated for the distribution of mycotoxin citrinin biosynthesis genes. These results showed that the genotype of citrinin genes was observed only in *M. purpureus* and *M. kaoliang*, while the phenotype of citrinin productivity was detected in *M. purpureus* and *M. kaoliang*. In contrast, the PCR and Southern blot results suggested that citrinin biosythesis genes were absent or significantly different in *M. pilosus*, *M. ruber*, *M. barkeri*, *M. floridanus*, *M. lunisporas*, and *M. pallens*. These results clearly indicated that the highly conserved citrinin gene cluster in *M. purpureus* and *M. kaoliang* carried out the citrinin biosynthesis. In addition, according to the phylogenetic subgroups established with the β-tubulin gene, the citrinin gene cluster can respectively group the species of *Monascus*.

Content

Characterization of MRT, a New Non-LTR Retrotransposon in *Monascus* spp.

Table

Figure

Overview

Chapter 1

Improving the Genetic Transformation of Filamentous Fungus *Monascus pilosus* Using Aurintricarboxylic Acid (ATA)

Chapter 2

Cloning and Characterization of Monacolin K Biosynthetic Gene Cluster from *Monascus pilosus*

Chapter 3

Overexpression of *mokA* Encoding the Polyketide Synthase of *Monascus pilosus* in *Escherichia coli*

Overview

Introdution

Monascus is a filamentous fungus which belongs to the class Ascomycetes and the family Monascaceae. Thirteen *Monascus* species have been reported. Asexual sporulation is a common reproductive mode for *Monascus* species. Sexual reproduction, which produces ascospores, has been observed (Carels and Shepherd 1975). *Monascus* species are known as producers of various secondary metabolites with polyketide structures as monacolins, pigments, and citrinin (Endo 1979; Shimizu et al., 2005). It has been applied in the food industry for thousands of years in Asia and discovered that *Monascus* produces several bioactive substances. These bioactive substances are mainly the secondary metabolites including substances for reducing hypertension, substances of anti-putrefaction bacteria such as monascidin, anti-cancer substances, substances for lowering blood sugar, ergosterol, anti-oxidants, and inhibitors of cholesterol synthesis such as monacolin. Moreover, the pigments produced by the various species of *Monascus* have been extensively used as natural food colorants (Babitha et al., 2007). *Monascus* can produce yellow, orange, and red pigments. The red pigment is regarded as the most important among these pigments because it can be used as substitutes for nitrites in meat products. It has enormous commercial value in terms of increasing the production of secondary metabolites. Therefore, *Monascus* has been valued as a functional health food.

Polyketide biosynthesis produced by microorganisms

Polyketide are secondary metabolites extensively produced by bacteria and fungi (Pfeifer and Khosla, 2001). The biosynthesis process of polyketide in

microorganisms is similar to bacterial and mammalian fatty acid synthesis. Bacteria produced polyketide are known to be classified into type I and type II polyketide synthase (PKS) systems according to the biosynthesis of polyketides (Pfeifer and Khosla, 2001). Modular system (type I PKS) contains multifunctional enzymes that are organized into modules. A module consisting of many motifs includes keto-synthase (KS), acyltransferase (AT), keto-reductase (KR), dehydratase (DH), enoyl rductase (ER), and acyl carrier protein (ACP) (Marsden et al., 1998) (**Figure 1**). Iterative system (type II PKS) contains each active sites encoded by separate enzymes. In addition, there is a significant characterization of type II polyketide compound with the aromatic group such as antibiotics doxorubicin (Pfeifer and Khosla, 2001) (**Figure 2**). However, the biosynthetic pathway of polyketides produced by fungi is different from bacteria. Fungal PKS defined as iterative type I PKS makes polyketides similar to bacterial type II PKS, but active sites of PKS are encoded by one enzyme instead of separate enzymes (Pfeifer and Khosla, 2001). According to the reduction of polyketide structures, fungal PKSs are classified into three types, highly-reduced (HR), partially-reduced (PR), and non-reduced (NR) (Nicholson et al., 2001). The structural diversity of polyketide also revealed the variety of biological activity. A number of polyketides have been used as drugs for human being diseases (Pfeifer and Khosla, 2001). For example, monacolin K produced by fungal iterative type I PKS has been used for reducing cholesterol synthesis.

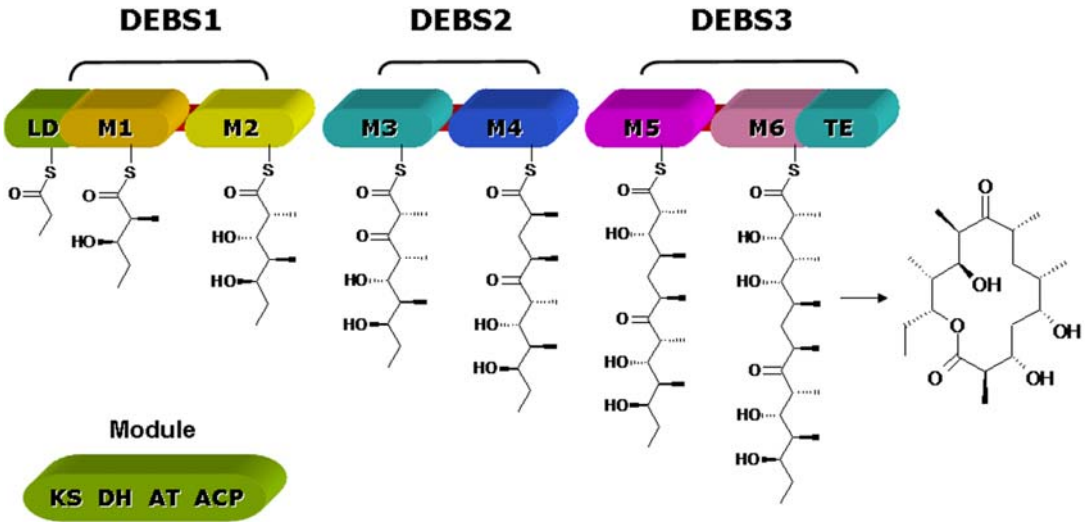


Figure 1. Biosynthesis of erythromycin from *Saccharopolyspora erythraea*.

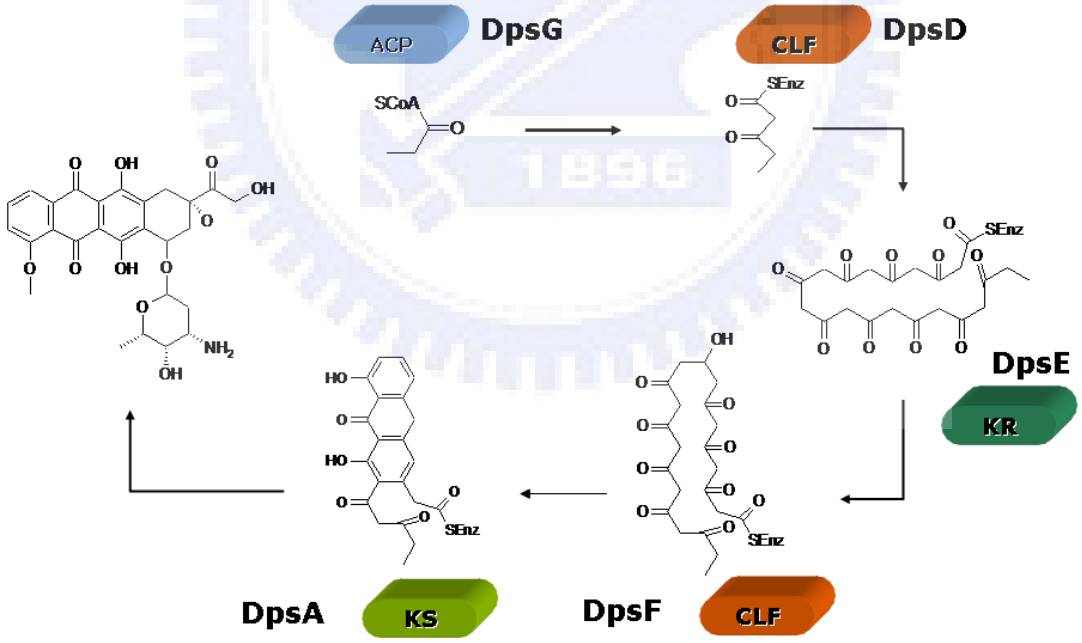


Figure 2. Biosynthesis of doxorubicin from *Streptomyce peucetius*.

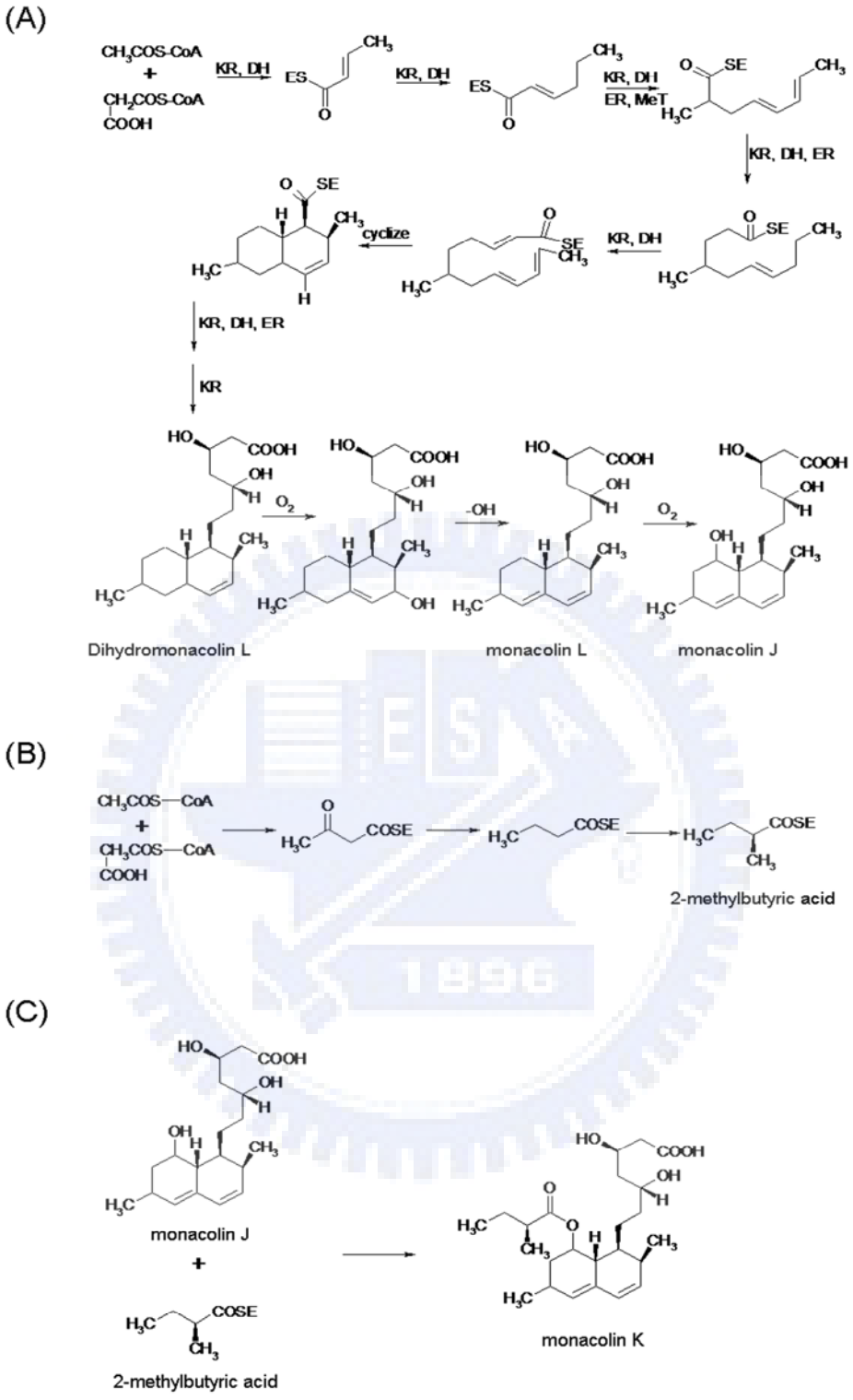
Monacolin K (Lovastatin) and Compactin biosynthesis

Monacolin K, the cholesterol synthesis inhibitor, is first isolated from the medium of *Monascus ruber* (Endo, 1979) and found the same substance from *Aspergillus terreus* as lovastatin (Hendrickson et al., 1999; Kennedy et al., 1999). The structure of monacolin K shares similarity with HMG-CoA; therefore, monacolin K competitively inhibits HMG-CoA reductase with HMG-CoA during cholesterol synthesis resulting in reduction of cholesterol synthesis. In previous studies, the lovastatin biosynthetic gene cluster has been proposed in *A. terreus* (**Figure 3**). Two polyketide synthases, transesterase, enoyl reductase, and P450 monooxygenase have been proved to involve in the structural biosynthesis of lovastatin (Hutchinson et al., 2000; Sorensen et al., 2003a; Sorensen et al., 2003b).

Figure 3. Lovastatin gene cluster from *Aspergillus terreus*. *lovB*, polyketide synthase; *lovF*, polyketide synthase; *lovA*, P450 monooxygenase; *lovG*, oxidoreductase; *lovC* dehydrogenase; *lovD*, transesterase; *lvrA*, HMG-CoA reductase; *lovE*, transcription factor and *lovI*, efflux pump.

The *lovB* encoding the polyketide synthase is responsible for nonaketide of lovastatin biosynthesis. According to the previous study, lovastatin can not be determined in the BX102 mutant, which is defective in *lovB* (Hendrickson et al., 1999). The cosmid library constructed from wild-type *A. terreus* is used to

complement the BX102 mutant. One transformant is identified with cosmid containing *lovB* gene for lovastatin producer (Hendrickson et al., 1999). The disruption of *lovC* blocks in lovastatin production and causes accumulation of the new compound, 4-Hydroxy-6-[(1*E*,3*E*,5*E*)-1-methylhepta-1,3,5-trien-1-yl]-2-pyrone. Coexpression of *lovB* and *lovC* in *A. nidulans* results in the production of dihydromonacolin L. Furthermore, the disruption of the *lovD* gene generates accumulation of monacolin J which is the precursor of lovastatin. Consequently, LovD could join together the two polyketide, 2-methylbutyryl and monacolin J, into lovastatin (**Figure 4**). The *lovE* encodes protein with binuclear Zn^{+2} finger motif characteristic of eukaryotic transcription factors (Kennedy et al., 1999). Transformation of an extra copy of the *lovE* gene-encoded transcription factor into the wild type strain resulted in a 7- to 10-fold overproduction of lovastatin (Hutchinson et al., 2000). The *lovI* mutant does not produce lovastatin or its known precursors. The deduced amino acid of *lovI* resembles known transmembrane metabolite transport proteins of microorganisms. In addition, *lovB* has successfully expressed in *Aspergillus nidulans*, and produces two novel polyketides, hexaketide pyrone (4-Hydroxy-6-[(1*E*,3*E*,5*E*)-1-methylhepta-1,3,5-trien-1-yl]-2-pyrone) and heptaketide pyrone (4-Hydroxy-6-[(1*E*,3*E*,5*E*,7*E*)-3-methylnona-1,3,5,7-tetraen-1-yl]-2-pyrone) (**Figure 5**).



Fugure 4. Lovastatin biosynthesis pathway from *A. terreus*. (A) *lovB* encodes nonaketide synthase is involved in the synthesis of monacolin J. (B) *lovF* encodes diketide synthase is involved in the synthesis of 2-methylbutyric acid. (C)

2-methylbutyryl and monacolin J join together into lovastatin by LovD.

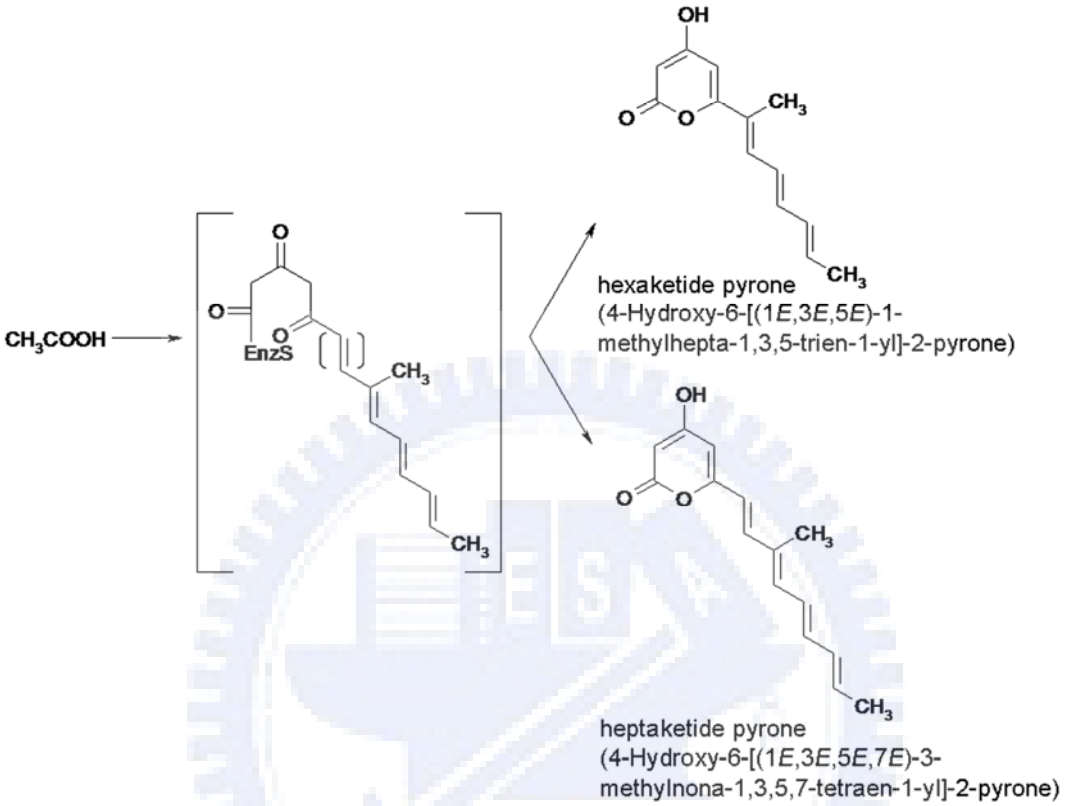


Figure 5. Hexaketide and heptaketide expressed in *A. nidulans* by harboring *lovB* gene from *A. terreus*.

Compactin is also the cholesterol synthesis inhibitor isolated from *Penicillium citrinum* (Abe et al., 2002). The structure of compactin differs from that of monacolin K, in which the C-6 position of the nonaketide-derived backbone lacks a methyl group derived from S-adenosyl-L-methionine (SAM). Recently, the compactin gene cluster is characterized (Abe et al., 2002a) (**Figure 6**). An extensive comparison analysis of these nine genes indicates compactin gene cluster

shared high similarity with the genes involved in lovastatin biosynthesis.

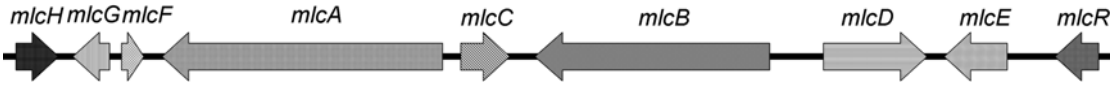


Figure 6. Compactin gene cluster from *Penicillium citrinum*. *mlcA*, polyketide synthase; *mlcB*, polyketide synthase; *mlcC*, P450 monooxygenase; *mlcF*, oxidoreductase; *mlcG* dehydrogenase; *mlcH*, transesterase; *mlcD*, HMG-CoA reductase; *mlcR*, transcription factor and *mlcE*, Efflux pump.

The polyketide synthases (*mlcA* and *mlcB*) and regulatory (*mlcR*) have been proved to be involved in compactin biosynthesis (Abe et al., 2002a; Abe et al., 2002b; Abe et al., 2002c) (**Figure 7**). The *mlcA* encoding polyketide synthase is responsible for nonaketide of compactin biosynthesis. As such, the disruption of *mlcA* results in the phenotype of lost compactin productivity (Abe et al., 2002a). The *mlcB* encoding polyketide synthase is responsible for diketide of compactin biosynthesis. The *mlcB* mutant demonstrates the accumulation of ML-236A (**Figure 7**). The *mlcR* that encodes the Zn(II)2Cys6 binuclear DNA binding protein is a regulatory of citrinin biosynthesis. Transformation of an extra copy of the *mlcR* gene-encoded transcription factor into the wild type strain results in a 10-15% increase of compactin (Abe et al., 2002b). Furthermore, introduction of the cosmid, which contains seven of the nine compactin biosynthetic genes, into *P. citrinum* can increase the yield of compactin of transformants (Abe et al., 2002c). It is interesting to compare the monacolin K, lovastatin, and compactin gene clusters from *Monascus*, *Aspergillus*, and *Penicillium*. They will provide useful

information to study evolution of secondary metabolite.

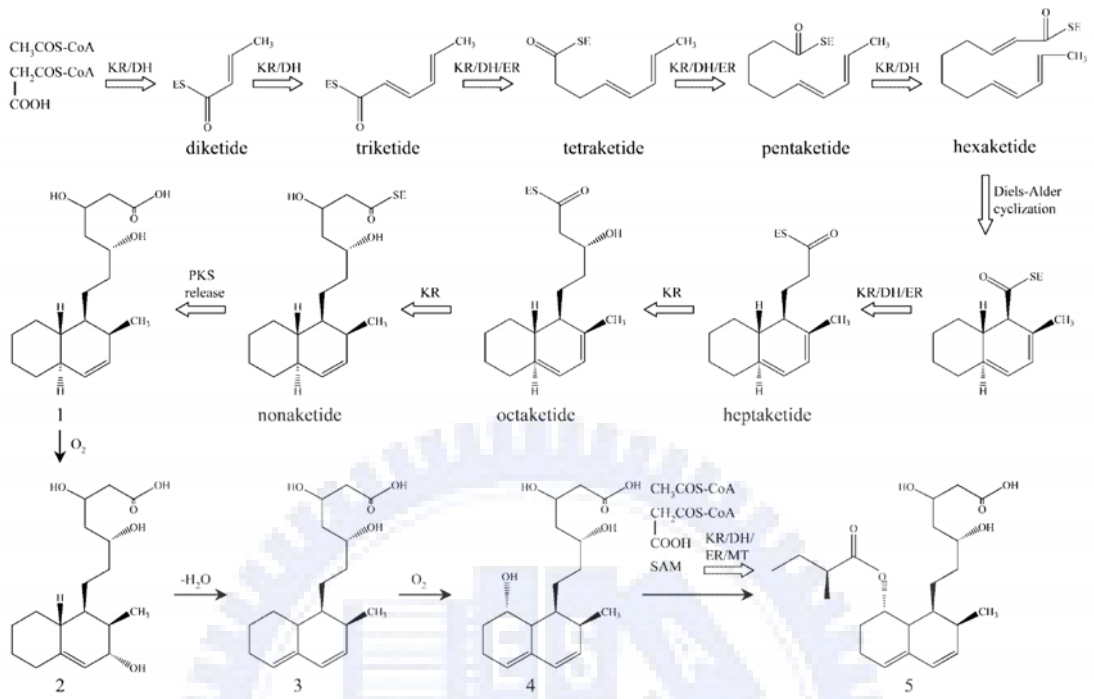


Figure 7. Compactin biosynthesis pathway from *P. citrinum*. 1: 4a,5-dihydro ML-236C; 2:3a-hydroxy-3,5-dihydro ML-236C, 3: ML-236C, 4: ML-236A; 5:ML-236B (compactin). SAM, S-adenosyl-L-methionine

Citrinin biosynthesis

Citrinin is a polyketide produced by fungal species including the genus *Aspergillus*, *Penicillium* and *Monascus*. It is identified as a contaminant in several foods and resulted in accumulation in the mitochondria inducing apoptosis at the cellular level. The biosynthesis of citrinin synthesized by the iterative type I polyketide synthase (PKS) originates from a tetraketide arising from the condensation of one acetyl-CoA molecule with three malonyl-CoA molecules in *Monascus* (Hajjaj et al., 1999) (**Figure 8**).

The citrinin biosynthetic gene cluster has been proposed in *M. purpureus* BCRC33325 (IFO30873) in Shimizu's studies (2005) (**Figure 9**). *pksCT* is predicted to be responsible for the synthesis of the polyketide skeleton. Also included are a dehydrogenase (*orf1*), a transcriptional activator (*ctnA*), an oxygenase (*orf3*), an oxidoreductase (*orf4*), and a transporter gene (*orf5*). The *pksCT* encoding polyketide synthase is responsible for such biosynthesis. Therefore, the disruption of *pksCT* results in the phenotype of lost citrinin productivity (Shimizu et al., 2005). The *ctnA* that encodes the Zn(II)2Cys6 binuclear DNA binding protein is a major activator of citrinin biosynthesis. The *ctnA*-disrupted strain of *M. purpureus* also brings a significant decrease in citrinin production down to a barely detectable level (Shimizu et al., 2007). Based on the citrinin gene cluster, we can characterize the distribution of citrinin in various *Monascus* species.

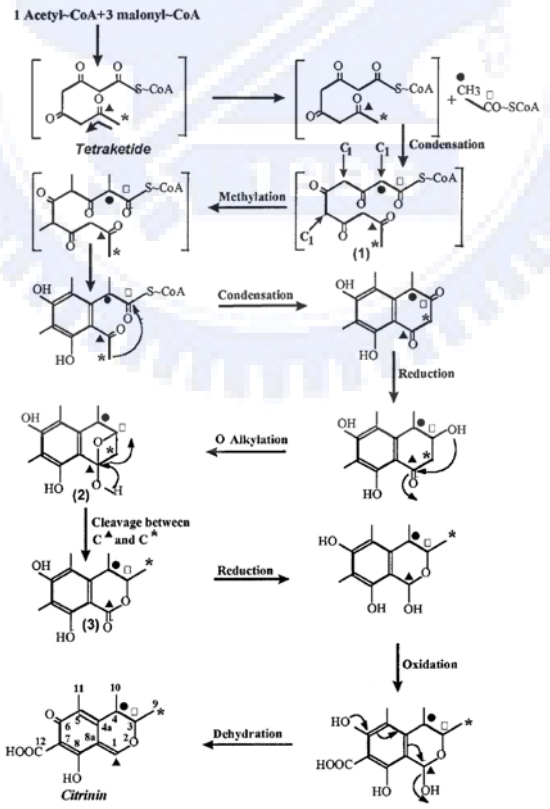


Figure 8. Citrinin biosynthesis pathway from *M. purpureus*.

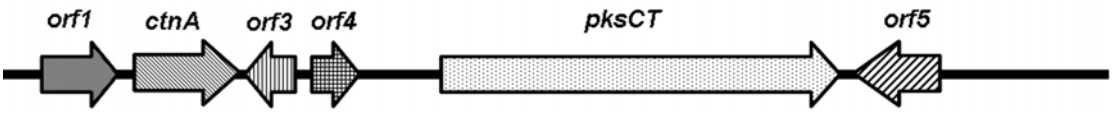


Figure 9. Citrinin gene cluster from *Monascus purpureus*. *orf1*, dehydrogenase; *ctnA*, transcriptional activator; *orf3*, oxygenase; *orf4*, oxidoreductase; *pksCT*, polyketide synthase; and *orf5*, transporter.

Retrotransposon

Retrotransposons are integral components of eukaryotic genome. The copy-and-paste mechanism of retrotransposons results in diversity of organisms and plays an important role in evolutionary history and population dynamics (Daboussi and Capy, 2003). Most retrotransposons are present in ancestral species and evolve with time. However, the distribution of retrotransposon is present in some species and lost in others, possibly due to genetic drift (Engels, 1981). Analyses of this distribution could exploit retrotransposons behavior in the various species, and allow us to group the species in the phylogenetic subgroups previously established with other characters (Cizeron et al. 1998; Blesa et al., 2001).

Retrotransposons can be divided into two major classes: long terminal repeat (LTR) and non-LTR retrotransposons. Approximately 30 transposable elements corresponding to both classes are recognized in fungi (Daboussi and Capy, 2003). Non-LTR retrotransposon, also known as long interspersed nuclear elements (LINE), contains a significant feature lacking in long terminal repeats. A mechanism of non-LTR retrotransposon for reverse transcription and integration of a new copy has been suggested in which reverse transcript becomes inserted at the nicks in the chromosome during DNA repair or recombination by host-encoded proteins (Flavell, 1995; Finnegan, 1997). Moreover, a relatively simple insertion mechanism of

non-LTR retrotransposon is proposed. Non-LTR retrotransposon inserts itself into the chromosome using the 3' end of the broken DNA digested by the endonuclease domain of retrotransposon's product to prime cDNA synthesis, a process termed target-primed reverse transcription (TPRT) (Chaboissier et al., 2000; Bibillo and Eickbush, 2002; Eickbush, 2002). Additionally, it has been suggested that DNA repair of a host's mechanism may involve the retrotransposition for the integration of the 5' end of *R2Bm* non-LTR retrotransposon (Fujimoto et al., 2004). Most non-LTR retrotransposons in fungi contain two open reading frames. The first one is considered to be the *gag* gene containing cysteine-rich nucleotide-binding domains though it does not show a conserved protein sequence. The second one encodes several conserved domains including endonuclease, reverse transcriptase, RNaseH domains, and zinc finger motifs (Daboussi and Capy, 2003). Extensive analysis of non-LTR retrotransposon in fungi has revealed the presence of numerous mutations resulting from the repeat-induced point mutation (RIP) process (Selker et al., 1987). The genome of *N. crassa* shows that most of the repetitive sequences (81%) have been mutated by RIP (Galagan et al., 2003). Methylation-induced premeiotically (MIP) occurs during the sexual phase of the fungal life cycle and leads to the methylation of cystosine residues (Faugeron et al., 1990). The occurrence of MIP also suggests that this process is directed against repetitive sequences (Goyon et al., 1996).

Five non-LTR retrotransposons have been characterized in fungi including *Tad1-1* in *Neurospora crassa* (Cambareri et al., 1994), MGR583 in *Magnaporthe grisea* (Hamer et al., 1989), *Cg*T1 in *Colletotrichum gloeosporioides* (He et al., 1996), *marY2N* in *Tricholoma matsutake* (Murata et al., 2001), and Mars in *Ascobollus immerses* (Goyon et al., 1996) though the entire element is not described for the last species. Non-LTR retrotransposons are classified into eleven clades

(Malik et al., 1999). Based on sequence, structure, and phylogenetic analyses, the elements from genomes of fungi belong to the Tad1 clade. Recently, two non-LTR retrotransposons of yeast, Zorro in *Candida albicans* (Goodwin et al., 2001) and Ylli in *Yarrowia lipolytica* (Casaregola et al., 2002), have been characterized. These two non-LTR retrotransposons belong to the L1 clade of elements known from human beings and mice. So far, no attempt was made to analyze the non-LTR retrotransposon in *Monascus*. It will provide useful information for the phylogenetic characterization in *Monascus*.

The specific aims of this study:

Monascus has been used in Chinese fermented foods for thousands of years. Moreover, many secondary metabolites have been identified such as monacolins, citrinin, GABA (gamma-aminobutyric acid), ergosterol, and various pigment compounds. However, only a few attempts have been made at genetic investigation of *Monascus* species. In the present study, a BAC (Bacterial Artificial Chromosome) library of the strain *M. pilosus* BCRC38072 about 12,000 clones was constructed for the genomic research. According to the BAC clones, monacolin K, citrinin, and non-LTR retrotransposon related genes were examined. The evolutionary relationship was further explored among different *Monascus* species. Therefore, the following specific aims were studied.

1. The efficiency of genetic transformation is depending on different *Monascus* strains. Although the genetic transformation of *M. ruber* and *M. purpureus* has been developed (Yang and Lee, 2008; Kim et al., 2003), no attempt has been made on genetic transformation of *M. pilosus* BCRC38072. Here, a method of genetic transformation of high efficiency was established using nuclease

inhibitor, aurintricarboxylic acid (ATA). The results demonstrated that ATA can improve the frequency of genetic transformation in *M. pilosus*. This valid and convenient method shows that ATA can be applied to fungi for efficient genetic transformation.

2. Although the lovastatin biosynthetic gene cluster in *A. terreus* has been characterized (Kennedy et al., 1999), the structural genes responsible for monacolin K (lovastatin) biosynthesis in *Monascus* are still unclear. Firstly, to explore the monacolin K biosynthetic gene cluster, a BAC library from *M. pilosus* BCRC38072 producing monacolin K was constructed (**Figure 10**). The nine genes were analyzed and found to share high similarity with the genes involved in the lovastatin biosynthetic gene cluster of *Aspergillus terreus* and compactin biosynthetic gene cluster of *Penicillium citrinum*. Moreover, the *mokA* encoding polyketide synthase and *mokH* encoding transcription factor were futher disrupted and overexpressed in *M. pilosus* BCRC38072 respectively.
3. Only a few attempts have been made at heterologous expression of fungal polyketide synthase. *mokA* encoding polyketide synthase was expressed in *Escherichia coli* for development of novel polyketide. However, an unexpected product, anthranilic acid which is the precursor of tryptophan, was observed. These results revealed that expression product of polyketide synthase (*mokA*) in *E. coli* was responsible for the elicitor to improve the activity of anthranilate synthase.
4. *Monascus* has been valued as a functional health food. Therefore, it is important to identify the non-citrinin producing *Monascus* strains or generally eliminate the production of citrinin in *Monascus*. Various *Monascus* species were analyzed for the presence of the citrinin genes and production of citrinin. The citrinin genes were observed only in *M. purpureus* and *M. kaoliang*.

Likewise, the citrinin productivity was detected in *M. purpureus* and *M. kaoliang*.

5. During the whole genome sequencing of *M. pilosus* BCRC38072, two repetitive sequences (mps01-1 and mps01-2) were observed in a *ca.* 160kb BAC, mps01 (**Figure 10**). These two repetitive sequences were new non-LTR retrotransposon named MRT (*Monascus* Retrotransposon). The entire nucleotide sequence of the MRT element was 5.5-kb long, including two open reading frames, ORF1 containing a cysteine-rich zinc finger motif and ORF2 containing apurinic/apyrimidinic endonuclease (APE), reverse transcriptase (RT), RNaseH domains, and a CCHC motif. Moreover, MRT elements were determined within *M. pilosus*, *M. ruber*, *M. sanguineus*, and *M. barkeri*.

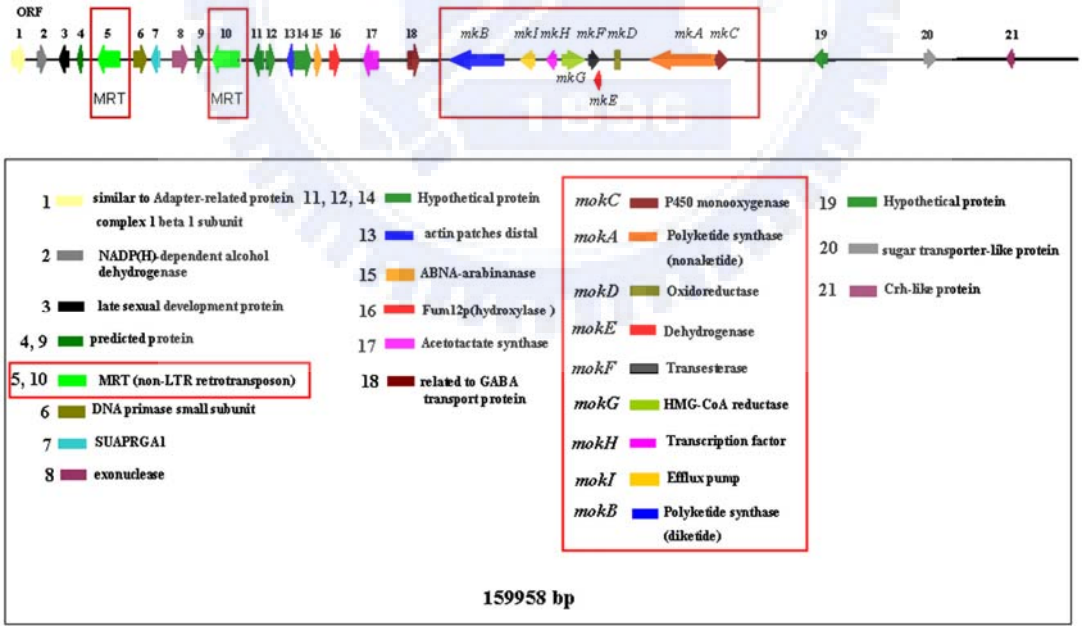


Figure 10. Annotation of BAC DNA mps01 predicted by BLAST and VectorNTI.

Reference

(1) Abe, Y.; Suzuki, T.; Ono, C.; Iwamoto, K.; Hosobuchi, M.; Yoshikawa, H. Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in *Penicillium citrinum*. *Mol. Genet. Genomics.* **2002a**, *267*, 636-646.

(2) Abe, Y.; Ono, C.; Hosobuchi, M.; Yoshikawa, H. Functional analysis of *mlcR*, a regulatory gene for ML-236B (compactin) biosynthesis in *Penicillium citrinum*. *Mol. Genet. Genomics.* **2002b**, *268*, 352-361

(3) Abe, Y.; Suzuki, T.; Mizuno, T.; Ono, C.; Iwamoto, K.; Hosobuchi, M.; Yoshikawa, H.; Effect of increased dosage of the ML-236B (compactin) biosynthetic gene cluster on ML-236B production in *Penicillium citrinum*. *Mol. Genet. Genomics.* **2002c**, *268*, 130-137.

(4) Babitha, S.; Soccol, C. R.; Pandey, A. Solid-state fermentation for the production of *Monascus* pigments from jackfruit seed. *Bioresour. Technol.* **2007**, *98*, 1554-1560.

(5) Bibillo, A.; Eickbush, T.H. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J. Mol. Biol.* **2002**, *316*, 459-473.

(6) Blesa, D.; Gandía, M.; Martínez-Sebastián, M.J. Distribution of the *bilbo* non-LTR retrotransposon in Drosophilidae and its evolution in the *Drosophila obscura* species group. *Mol. Biol. Evol.* **2001**, *18*, 585-592.

(7) Cambareri, E.B.; Helber, J.; Kinsey, J.A. *Tad1-1*, an active LINE-like element of *Neurospora crassa*. *Mol. Gen. Genet.* **1994**, *242*, 658-665.

(8) Carels, M.; Shepherd, D. Sexal reproductive cycle of *Monascus* in submerged shaken culture. *J. Bacteriol.* **1975**, *122*, 288-294.

(9) Casaregola, S.; Neuvéglise, C.; Bon, E.; Gaillardin, C. Ylli, a non-LTR

retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. *Mol. Biol. Evol.* **2002**, *19*, 664-677.

(10) Chaboissier, M.C., Finnegan, D.; Bucheton, A. Retrotransposition of the I factor, a non-long terminal repeat retrotransposon of *Drosophila*, generates tandem repeats at the 3' end. *Nucleic Acids Res.* **2000**, *28*, 2467-2472.

(11) Cizeron, G.; Lemeunier, F.; Loevenbruck, C.; Brehm, A.; Biémont., C. Distribution of the retrotransposable element 412 in *Drosophila* species. *Mol. Biol. Evol.* **1998**, *15*, 1589-1599.

(12) Daboussi, M.J.; Capy, P. Transposable elements in filamentous fungi. *Annu. Rev. Microbiol.* **2003**, 57, 275-299.

(13) Eickbush, T.H. Repair by retrotransposition. *Nat. Genet.* **2002**, *31*, 126-127.

(14) Endo, A. Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* (Tokyo) **1979**, *32*, 852-854.

(15) Engels, W.R. Estimating genetic divergence and genetic variability with restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* **1981**, *78*, 6329-6333.

(16) Faugeron, G., Rhounim, L.; Rossignol, J.L. How does the cell count the number of ectopic copies of a gene in the premeiotic inactivation process acting in *Ascobolus immersus*? *Genetics*. **1990**, *124*, 585-591.

(17) Finnegan, D.J. Transposable elements: how non-LTR retrotransposons do it. *Curr. Biol.* **1997**, *7*, 245-248.

(18) Flavell, A.J. Retroelements, reverse transcriptase and evolution. *Comp. Biochem. Physiol.* **1995**, *110*, 3-15.

(19) Fujimoto, H.; Hirukawa, Y.; Tani, H.; Matsuura, Y.; Hashido, K.; Tsuchida, K.; Takada, N.; Kobayashi, M.; Maekawa, H. Integration of the 5' end of the retrotransposon, R2Bm, can be complemented by homologous recombination. *Nucleic Acids Res.* **2004**, *32*, 1555-1565.

(20) Galagan, J.E.; Calvo, S.E. et al. (77 co-authors). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. **2003**, *422*, 859-868.

(21) Goodwin, T.J.; Ormandy, J.E.; Poulter, R.T. L1-like non-LTR retrotransposons in the yeast *Candida albicans*. *Curr. Genet.* **2001**, *39*, 83-91.

(22) Goyon, C., Rossignol, J.L.; Faugeron, G. Native DNA repeats and methylation in *Ascobolus*. *Nucleic. Acids. Res.* **1996**, 24, 3348-3356.

(23) Hajjaj, H.; Klaebe, A.; Loret, M. O.; Goma, G.; Blanc, P. J.; Francois, J. Biosynthetic pathway of citrinin in the filamentous fungus *Monascus ruber* as revealed by ^{13}C nuclear magnetic resonance. *Appl. Environ. Microbiol.* **1999**, *65*, 311-314.

(24) Hamer, J.E.; Farrall, L.; Orbach, M.J.; Valent, B.; Chumley, F.G. Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. *Proc. Natl. Acad. Sci. USA.* **1989**, *86*, 9981-9985.

(25) He, C.; Nourse, J.P.; Kelemu, S.; Irwin, J.A.; Manners, J.M. *Cg*T1: a non-LTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. *Mol. Gen. Genet.* **1996**, *252*, 320-331.

(26) Hendrickson, L.; Davis, C. R.; Roach, C.; Nguyen, D. K.; Aldrich, T.; McAda, P. C.; Reeves, C. D. Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem. Biol.* **1999**, *6*, 429-439.

(27) Hutchinson, C. R.; Kennedy, J.; Park, C.; Kendrew, S.; Auclair, K.; Vederas, J. Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. *Antonie Van Leeuwenhoek.* **2000**, *78*, 287-295.

(28) Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R. Modulation of polyketide synthase activity by accessory proteins during

lovastatin biosynthesis. *Science.* **1999**, *284*, 1368-72.

(29) Kim, J.G.; Choi, Y.D.; Chang, Y.J.; Kim, S.U. Genetic transformation of *Monascus purpureus* DSM1379. *Biotechnol. Lett.* **2003**, *25*, 1509-1514.

(30) Malik, H.S; Burke, W.D.; Eickbush, T.H. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **1999**, *16*, 793-805.

(31) Marsden, A.F.; Wilkinson, B.; Cortés, J.; Dunster, N.J.; Staunton, J.; Leadlay, P.F. Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science*. **1998**, *279*, 199-202.

(32) Murata, H.; Miyazaki, Y.; Yamada, A. *marY2N*, a LINE-like non-long terminal repeat (non-LTR) retroelement from the ectomycorrhizal homobasidiomycete *Tricholoma matsutake*. *Biosci. Biotechnol. Biochem.* **2001**, *65*, 2301-2305.

(33) Nicholson, T.P.; Rudd, B.A.; Dawson, M.; Lazarus, C.M.; Simpson, T.J.; Cox, R.J. Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem. Biol.* **2001**, *8*, 157-178.

(34) Pfeifer, B. A.; Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 106-118.

(35) Selker, E.U.; Cambareri, E.B.; Jensen, B.C.; Haack, K.R. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell*. **1987**, *51*, 741-752.

(36) Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T. Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2005**, *71*, 3453-3457.

(37) Shimizu, T.; Kinoshita, H.; Nihira, T. Identification and in vivo functional analysis by gene disruption of *ctnA*, an activator gene involved in citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2007**, *73*, 5097-5103.

(38) Sorensen, J. L.; Auclair, K.; Kennedy, J.; Hutchinson, C. R.; Vederas, J. C.

Transformations of cyclic nonaketides by *Aspergillus terreus* mutants blocked for lovastatin biosynthesis at the *lovA* and *lovC* genes. *Org. Biomol. Chem.* **2003a**, *1*, 50-59.

(39) Sorensen, J. L.; Vederas, J. C. Monacolin N, a compound resulting from derailment of type I iterative polyketide synthase function *en route* to lovastatin. *Chem. Commun.* **2003b**, *13*, 1492-1493.

(40) Yang, Y.J.; Lee, I. Agrobactrium tumefaciens-mediated transformation of *Monascus ruber*. *J. Microbiol. Biotechnol.* **2008**, *18*, 754-758.

Chapter 1

Improving the Genetic Transformation of Filamentous Fungus *Monascus pilosus* Using Aurintricarboxylic Acid (ATA)

Abstract

The filamentous fungus, *Monascus pilosus*, was genetically transformed with a reporter plasmid, pMS-1.5hp, using aurintricarboxylic acid (ATA) treatment to obtain a high red pigment-producing mutant. The transformation efficiency of *Monascus pilosus* was higher with the ATA-treatment than with either a non-restriction-enzyme-mediated integration (REMI) or a REMI method. This valid and convenient random mutagenesis method shows that ATA can be applied in fungi for efficient genetic transformation.

Key words: *Monascus pilosus*; restriction-enzyme-mediated integration (REMI); aurintricarboxylic acid (ATA).

Introduction

Monascus spp. are filamentous fungi that are known as producers of various secondary metabolites with polyketide structures as pigments, monacolins and citrinin (Endo et al., 1986; Shimizu et al., 2005). The production of red pigments by various species of *Monascus* has been extensively used as natural food colorants (Babitha et al., 2007a). Monacolin K, also known as lovastatin for reducing serum cholesterol levels in human beings, was first isolated from the medium of *Monascus ruber* (Endo et al., 1979; Tobert, 2003). The development of the transformation system of *Monascus* should enable the function of polyketide biosynthesis to be examined by constructing knock-out mutants, or high-economic polyketides, such as monacolins and pigments, to be produced.

The preparation of fungal protoplasts is usually the typical approach for delivering the transforming DNA into filamentous fungus. The protoplasts are mixed with a combination of $CaCl_2$ and polyethylene glycol (PEG), which cultivate the transforming DNA. Restriction-enzyme-mediated integration (REMI) (Kahmann and Basse, 1999) and the *Agrobacterium tumefaciens*-mediated transformation (ATMT) (de Groot et al., 1998) methods have been established to improve the frequency of transformation substantially. The REMI method has been adopted in various filamentous fungi (Cantone and Vandenberg, 1999; Mullins and Kang, 2001; Thon et al., 2000). The method is based on the addition of a restriction enzyme into the mix of protoplasts and the transforming DNA, such that the specific restriction enzyme generates a site for inserting DNA into genomic locus (Mullins and Kang, 2001). The process increases the efficiency of transformation; however, adding different restriction enzymes and concentrations yields different efficiencies of transformation (Cantone and Vandenberg. 1999; Jin et al., 2005).

Aurintricarboxylic acid (ATA) is an inhibitor for the inhibition of the activity of DNaseI, S1 nuclease, exonuclease III and RNAse A (Hallick et al., 1977). It has been demonstrated to be useful in protecting plasmid DNA from nuclease degradation and increasing the transfer efficiency of DNA in the respiratory tissues and the skin of mice, pigs and macaques (Glasspool-Malone and Malone, 1999; Glasspool-Malone et al., 2000; Glasspool-Malone et al., 2002; Walther et al., 2005). In this study, the use of ATA was investigated for the genetic transformation in *M. pilosus*. The combination of $CaCl_2$ and PEG was employed to analyze the efficiency of genetic transformation between ATA-treatment and REMI. The transformation plasmid, including the promoter of heat shock protein 90 (*hsp* 90) from *M. pilosus*, was constructed. Additionally, the fusion of the hygromycin B resistance gene (HPH) with enhanced green fluorescent protein (EGFP) was expressed in *M. pilosus*.

Materials and methods

Strain used and growth conditions

M. pilosus BCRC38072, a monacolin K-producing strain that was isolated from red rice (anka) collected from a traditional local market, was used in this study. To prepare the protoplasts, the strain was incubated on PDB (DIFCO 254920, Detroit, Michigan) agar for one week, and spore suspensions were obtained by washing cultured PDA plates.

Construction of transformation vector

The human cytomegalovirus (CMV) promoter of plasmid pHygEGFP (BD

Biosciences Clontech, Palo. Alto, CA) was digested by *Bgl*II-*Xho*I and blunted using an End-It DNA End-repair kit (Epicentre, Madison, WS) to obtain the linear plasmid. The heat shock protein 90 (*hsp* 90) promoter from *M. pilosus* BCRC38072 (GenBank accession no. DQ983312) was amplified by the primer set (Mps9259F: AGTGGCAGCCAACCCTCACC and Mps7782R: CGGGCTGATAGAGCAGATAGATAGATG). The 1.5kb PCR product was phosphorylated and introduced into the linear plasmid pHygEGFP to obtain the plasmid pMS-1.5hp.

Protoplasts isolation

The preparation of protoplast developed from *Neurospora crassa* (Vollmer and Yanofsky, 1986) was modified for the genetic transformation of *M. pilosus*. The conidia from the one-week culture of *M. pilosus* were incubated in 100 ml of Vogel medium at 30^{o}C for 16~18 hr. The mycelia were harvested on miracloth (Millipore, MA, USA) and washed in the MA digestion solution (0.1 M maleic acid, pH 5.5 and 1.2 M $(NH_4)_2SO_4$). The mycelia were digested for 4~5 hr using 100 mg of Yatalase (Takara, Shiga, Japan), 100 mg of lysing enzyme (Sigma, St.Louis, Mo., USA) and 100 μl of β-glucuronidase (Sigma, St.Louis, Mo., USA) in 50 ml MA digestion solution. To remove the undigested mycelia, protoplasts were harvested by passing them through miracloth and by centrifugation at 1000 rpm for 10 min (Sorvall, CT, USA). The protoplasts were maintained in 80 % STC (1 M sorbitol, 50 mM Tris pH 8.0, 50 mM $CaCl_2$) and 20 % PTC (40 % PEG 4000, 50 mM Tris pH 8.0, 50 mM $CaCl_2$), and dimethylsulfoxide (DMSO) was added to a final concentration of 1 %.

Transformation of *M. pilosus* BCRC38072

For genetic transformation of *M. pilosus*, the 100 μl protoplasts were mixed

with 5 μg DNA-linearized and ATA (1 mM) or restriction enzyme (10 units). The mixtures were incubated on ice for 30 min. One milliliter PTC was added and mixed gently. Following incubation at room temperature for 20 min, the protoplast mixtures were added to 15 ml of SYP medium (1 M sorbitol, 0.1 % yeast extract, 0.1 % peptone and 2 % agar) that contained 60 μg hygromycin B/ml. The transformants were incubated at 30 ^{o}C.

Manipulations of nucleic acid

Fungal genomic DNA was isolated by liquid nitrogen treatment according to the method developed by Bingle et al. (1999). Southern hybridization was performed using the DIG system (Roche Diagnostics, Mannheim, Germany). The probe of *hph* gene was DIG-labeled by PCR amplification using the PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The primer set of *hph* gene was pMS-f: GAACTGCCCGCTGTTCTGCA and pMS-r: ACATCGCCTCGCTCCAGTCA.

Fluorescence microscopy

The hyphae of transformants were compared by fluorescence microscopy to that of the wild type in EGFP expression. Fluorescence photomicrography of the specimens was conducted using a fluorescence microscope (Leica RXA, Mannheim, Germany). The red filter (Chroma, VT, USA) with peak transmission of excitation at 450 to 490 nm and peak transmission of suppression at 525 to 575 nm was used to detect EGFP.

Measurement of red pigment

To detect the red pigment, the strains were incubated on PDA agar for one week,

and spore suspensions were obtained by washing cultured PDA agar plates with distilled water. Mycelia and media were harvested after incubating 3~10 days at 25 ^{o}C with constant agitation in liquid medium (7 % glycerol, 3 % glucose, 3 % monosodium glutamate, 1.2 % polypetone, 0.2 % $NaNO_3$ and 0.1 % $MgSO_4.7H_2O$). The aliquots of *M. pilosus* and the transformant culture were cleared of cells, and were filtered by a 0.2 mm filter. The supernatants were analyzed at 500 nm by the spectrophotometer for the determination of red pigment.

Results and discussion

Preparation of protoplasts and sensitivity test of *M. pilosus* BCRC38072 to hygromycin B

The mycelia of *M. pilosus* BCRC38072 were initially white, and then became reddish orange. The size of conidia was 10~13 × 8~10 μm and the ascospore was ellipsoid with dimensions of 4.6~6.3 × 3.3~4.2 μm. The conidia obtained after cultivation for seven days was transferred into the Vogel medium and was over 90 % germinated at 30 ^{o}C for 16~18 hr. The result revealed that a mixture of 100 mg Yatalase, 100 mg lysing enzyme and 100 μl β-glucuronidase was sufficient to convert hyphae (**Figure 1A**) into protoplasts (**Figure 1B**) when digested for 4~5 h in 50 ml of digestion solution, after which the protoplasts were used for genetic transformation. The protoplast yield for *M. pilosus* was $1.0\sim3.0\times10^{7}$ ml^{-1}.

The inhibition of the growth of *M. pilosus* BCRC38072 was determined by plating conidia on a PDA plate with various hygromycin B concentrations - 0, 20, 30, and 40 μg/ml. The growth of *M. pilosus* BCRC38072 was completely inhibited at 30 μg/ml. However, the false-positive colonies were commonly involved in the

genetic transformation of fungi (Mort-Bontemps and Fèvre, 1997; Shimizu et al., 2006). To ensure the absence of false-positive colonies, high concentration of hygromycin B (60 μg/ml) was used to select resistant colonies in genetic transformation experiments.

Evaluation of efficiency of genetic transformation using the expression vector pMS-1.5hp

Hsp90, the 90kDa heat shock protein, is a major molecular chaperon which is distributed in all eukaryotes. It is the most abundant protein and occupies 1~2 % of all cellular proteins in the cytosols of eukaryotic cells (Csermely et al., 1998). Therefore, the construction of the expression vector using the *hsp*90 promoter from *M. pilosus* may promote the production of hygromycin-resistance protein. In this study, based on the Bacterial Artificial Chromosome (BAC) library constructed from *M. pilosus* BCRC38072 (Chen et al., 2007), the primer set was designed to amplify the promoter of the *hsp*90 gene (GenBank accession no. DQ983312). The human cytomegalovirus (CMV) promoter of the expression vector pHygEGFP, including a fusion of the hygromycin B resistance gene (HPH) with enhanced green fluorescent protein (EGFP), was replaced with the promoter of the *hsp*90 gene from *M. pilosus* BCRC38072 to yield the plasmid pMS-1.5hp (**Figure 2A**). Furthermore, a time course of pMS-1.5hp plasmid stability was analyzed in lysates from *M. pilosus* pretreated with 0, 0.1 and 1.0 mM ATA. We observed that 0.1 mM ATA protected the plasmid in 1 mg ml^{-1} of protein lysate, whereas the stability of the plasmid was decreased in 3.5 mg ml^{-1} of protein lysate (**Figure 2B, C**). These results indicate that the presence of higher ATA concentrations (1.0 mM) can prolong the stability of plasmid DNA. The expression vector was utilized to achieve the over-expression of the dual-function marker for the genetic transformation in *M. pilosus* BCRC38072.

The REMI method was established to substantially improve the frequency of transformation and adopted for use in various filamentous fungi (Cantone and Vandenberg, 1999; Mullins and Kang, 2001; Thon et al., 2000). The REMI method increases the efficiency of transformation; however, different efficiencies of transformation are achieved by adding different restriction enzymes at different enzyme concentrations (Cantone and Vandenberg, 1999; Jin et al., 2005). To understand the efficiency of genetic transformation with ATA treatment (1.0 mM), three procedures, including non-REMI, REMI and ATA-treatment, were carried out using the PEG-based protoplasts method (Vollmer and Yanofsky, 1986). The non-REMI method using *Fsp*I-linearized pMS-1.5hp had the lowest efficiency of genetic transformation (17 ± 2 transformants per 5 μg of linear plasmid), while the *Fsp*I-REMI had an efficiency of genetic transformation (48 ± 4 transformants per 5 μg of linear plasmid) that was 2.8 times greater than the non-REMI treatment (**Table 1**). The *Hpa*I-REMI had a seven times higher efficiency of genetic transformation (119 ± 11 transformants per 5 μg of linear plasmid) than the non-REMI treatment. Interestingly, the ATA-treatment yielded a 9.4 times higher efficiency of genetic transformation (159 ± 24 transformants per 5 μg of linear plasmid) when compared to the non-REMI transformation and a 3.3 and 1.3 times higher efficiency of genetic transformation compared to the *Fsp*I-REMI and *Hpa*I-REMI respectively. These results suggest that adding ATA to the protoplasts increases the efficiency of genetic transformation in *M. pilosus* BCRC38072. This result is consistent with the results reported for the transformation of the respiratory tissues and skin of mice, pigs and macaques (Glasspool-Malone and Malone, 1999; Glasspool-Malone and Malone, 2002; Glasspool-Malone et al., 2000; Glasspool-Malone et al., 2002; Walther et al., 2005). The ATA effect on genetic transformation is likely a result of ATA's ability to significantly reduce nuclease activity in cells, which prevents the transforming DNA from being degraded (Glasspool-Malone et al., 2002). The results of this study are also consistent with the findings of Cantone and Vandenberg, who

indicated that adding ATA results in a high efficiency of genetic transformation in the pathogen *Paecilomyces fumosoroseus* (Cantone and Vandenberg, 1999).

Characterization of the high red pigment transformant

Monascus produces various pigments that have enormous commercial value in terms of increasing the production of pigments (Babitha et al., 2007a). Previous studies have shown that high temperature, high concentration of NaCl, and improvement in oxygen supply can increase the yield of pigments (Babitha et al., 2007b; Hajjaj et al., 1999). In this study, 30 transformants were screened and one of these that had the highest production of red pigment was characterized. The result showed that random mutagenesis with efficient genetic transformation using ATA can obtain the transformant with high red pigment. The transformant T7 was compared with the wild type during submerged fermentation of 10 days. The red pigment production of the T7 transformant was 4 to 5-fold higher than the corresponding values of the wild type, while the cultured cell mass revealed no significant difference (**Figure 3**). After 5 days of cultivation, the red pigment increased in the T7 transformant. The maximum red pigment level was reached after 8 days of cultivation; however, the red pigment was slowly produced by the wild type and reached the maximum level during the 10th day of cultivation.

Confirmation of *M. pilosus* transformants

The high red pigment-producing transformant and seven other transformants of *M. pilosus* BCRC38072 were selected for Southern hybridization and fluorescence microscopy analyses. Their identities were also verified by transferring the stably grown colonies onto the new plates with hygromycin. The eight transformants were identified by Southern hybridization using *Hin*dIII and *Pvu*II restriction

enzymes (**Figure 4A, B**). *Hin*dIII cuts at one site in the pMS-1.5hp plasmid, while *Pvu*II does not cut within the plasmid. The results of Southern hybridization showed that the T5 transformant had a single copy integration, while T6 and T8 revealed the two integration events (two copies). Using the *Hin*dIII restriction enzyme, 6kb and 10kb fragments were observed in the transformants labeled as T1, T3 and T4, while 4kb and 6kb fragments corresponded to the transformants that were labeled as T2 and T7 (**Figure 4A**). However, only one fragment (>10 kb), corresponding to the transformants T1, T2, T3, T4 and T7, was detected when *Pvu*II was used as the restriction enzyme (**Figure 4B**). These results suggested that the plasmid had integrated as a tandem repeat in the transformants T1, T2, T3, T4 and T7. They were attributed to the rearrangement of the integrated plasmid. Further studies on the loci of gene disruption and the mechanism of gene regulation will clarify the molecular mechanisms by which the biosynthesis of red pigment occurs. To confirm that the transformants expressed EGFP from the integrated DNA, the expression of EGFP was detected (**Figure 4C, D**) using a fluorescence microscopy. Fluorescence photomicrography of the specimens was conducted and fluorescence was clearly observed in the transformants when the samples were excited with a red filter. The red filter, with a peak transmission of excitation at 450 to 490 nm and a peak transmission of suppression at 525 to 575 nm, was used to detect EGFP. These results indicated that the EGFP protein was extensively expressed in the mycelia and spores. Moreover, the highest red pigment-producing mutation was stably inherited through the next generation after several rounds of cultivation.

In conclusion, this study demonstrates that the nuclease inhibitor aurintricarboxylic acid (ATA) can be adopted to stably transform *M. pilosus*. As was shown, the addition of ATA increased the efficiency of genetic transformation in *M. pilosus*. A major advantage of ATA is that different restriction enzymes need not be tested. Due to the fact that the transformant with high red pigment can be isolated by random mutagenesis with efficient genetic transformation using ATA,

this valid and convenient method could be used to screen the transformant with high monacolin K, as well as other high-economic polyketides.

References

(1) Babitha, S.; Soccol, C.R.; Pandey, A. Solid-state fermentation for the production of *Monascus* pigments from jackfruit seed. *Bioresour. Technol.* **2007a**, *98*, 1554-1560.

(2) Babitha, S.; Soccol, C.R.; Pandey, A. Effect of stress on growth, pigment production and morphology of *Monascus* sp. in solid cultures. *J. Basic Microbiol.* **2007b**, *47*, 118-126.

(3) Bingle, L.E.H.; Simpson, T.J.; Lazarus, C.M. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet. Biol.* **1999**, *26*, 209-223.

(4) Cantone, F.A.; Vandenberg, J.D. Genetic Transformation and Mutagenesis of the Entomopathogenic Fungus *Paecilomyces fumosoroseus*. *J. Invertebr. Pathol.* **1999**, *74*, 281-288.

(5) Chen, Y.-P.; Tseng, C.-P.; Liaw, L.-L.; Wang, C.-L.; Yuan, G.-F. Characterization of MRT, a New Non-LTR Retrotransposon in *Monascus* spp. *Bot. Bull. Acad. Sin.* **2007**, *48*, 377-385.

(6) Csermely, P.; Schnaider, T.; Sőti, C.; Prohászka, Z.; Nardai, G. The 90-kDa molecular chaperone family: structure, function, and clinical applications. A comprehensive review. *Pharmacol. Ther.* **1998**, *79*, 129-168.

(7) de Groot, M.J.; Bundock, P.; Hooykaas, P.J.; Beijersbergen, A.G. Agrobacterium tumefaciens-mediated transformation of filamentous fungi. *Nat. Biotechnol.* **1998**, *16*, 839-842.

(8) Endo, A. Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* (Tokyo) **1979**, *32*, 852-854.

(9) Glasspool-Malone, J.; Malone, R.W. Marked enhancement of direct respiratory

tissue transfection by aurintricarboxylic acid. *Hum. Gene. Ther.* **1999**, *10*, 1703-1713.

(10) Glasspool-Malone, J.; Malone, R.W. Enhancing direct in vivo transfection with nuclease inibitors and pulsed electrical fields. *Methods Enzymol.* **2002**, *346*, 72-91.

(11) Glasspool-Malone, J.; Somiari, S.; Drabick, J.J.; Malone, R.W. Efficient nonviral cutaneous transfection. *Mol. Ther.* **2000**, *2*, 140-146.

(12) Glasspool-Malone, J.; Steenland, P.R.; McDonald, R.J.; Sanchez, R.A.; Watts, T.L.; Zabner, J.; Malone, R.W. DNA transfection of macaque and murine respiratory tissue is greatly enhanced by use of nuclease inhibitor. *J. Gene. Med.* **2002**, *4*, 323-332.

(13) Hajjaj, H.; Blanc, P.J.; Groussac, E.; Goma, G.; Uribelarrea, J.L.; Loubiere, P. Improvement of red pigment/citrinin production ratio as a function of environmental conditions by *Monascus ruber*. *Biotechnol. Bioeng.* **1999**, *64*, 497-501.

(14) Hallick, R.B.; Chelm, B.K.; Gray, P.W.; Orozco, E.M., Jr. Use of aurintricarboxylic acid as an inhibitor of nuclease during nucleic acid isolation. *Nucleic Acids Res.* **1977**, *4*, 3055-3064.

(15) Jin, X.; Mo, M.-H.; Wei, Z.; Huang, X.W.; Zhang, K.Q. Transformation and mutagenesis of the nematode-trapping fungus *Monacrosporium sphaeroides* by restriction enzyme-mediated integration (REMI). *J. Microbiol.* **2005**, *43*, 417-423.

(16) Kahmann, R.; Basse, C. REMI (Restriction Enzyme Mediated Integration) and its impact on the isolation of pathogenicity genes in fungi attacking plants. *Eur. J. Plant Pathol.* **1999**, *105*, 221-229.

(17) Mort-Bontemps, M.; Fèvre, M. Transformation of the oomycete Saprolegnia

monoïca to hygromycin-B resistance. *Curr. Genet.* **1997**, *31*, 272-275.

(18) Mullins, E.D.; Kang, S. Transformation: a tool for studying fungal pathogens of plants. *Cell. Mol. Life Sci.* **2001**, *58*, 2043-2052.

(19) Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T. Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2005**, *71*, 3453-3457.

(20) Shimizu, T.; Kinoshita, H.; Nihira, T. Development of transformation system in *Monascus purpureus* using an autonomous replication vector with aureobasidin A resistance gene. *Biotechnol. Lett.* **2006**, *28*, 115-120.

(21) Thon, M.R.; Nuckles, E.M.; Vaillancourt, L.J. Restriction enzyme-mediated integration used to produce pathogenicity mutants of *Colletotrichum graminicola*. *Mol. Plant-Microbe Interact.* **2000**, *13*, 1356-1365.

(22) Tobert, J.A. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* **2003**, *2*, 517-526.

(23) Vollmer, S.J.; Yanofsky, C. Efficient cloning of genes of *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA.* **1986**, *83*, 4869-4873.

(24) Walther, W.; Stein, U.; Siegel, R.; Fichtner, I.; Schlag, P.M. Use of the nuclease inhibitor aurintricarboxylic acid (ATA) for improved non-viral intratumoral in vivo gene transfer by jet-injection. *J. Gene Med.* **2005**, *7*, 477-485.

Table 1. The number of transformants based on various transformation methods. The plasmid pMS-1.5hp was used for genetic transformation.

Experiment	Transformation methods			
	Linear plasmid (non-REMI)[a]	Treated by ATA (1mM)[b]	REMI	
			*Fsp*I (10 units)[c]	*Hpa*I (10 units)[d]
M. pilosus[e]	17 ± 2	159 ± 24	48 ± 4	119 ± 11

a. The plasmid pMS-1.5hp was linearized by *Fsp*I restriction enzyme.

b. The *Fsp*I-linearized pMS-1.5hp was mixed with 1 mM ATA and protoplasts.

c. The *Fsp*I-linearized pMS-1.5hp was mixed with 10 units *Fsp*I restriction enzyme and protoplasts.

d. The *Hpa*I-linearized pMS-1.5hp was mixed with 10 units *Hpa*I restriction enzyme and protoplasts.

e. Number of transformants per 5 ug linear plasmid. The values were based on three determinations.

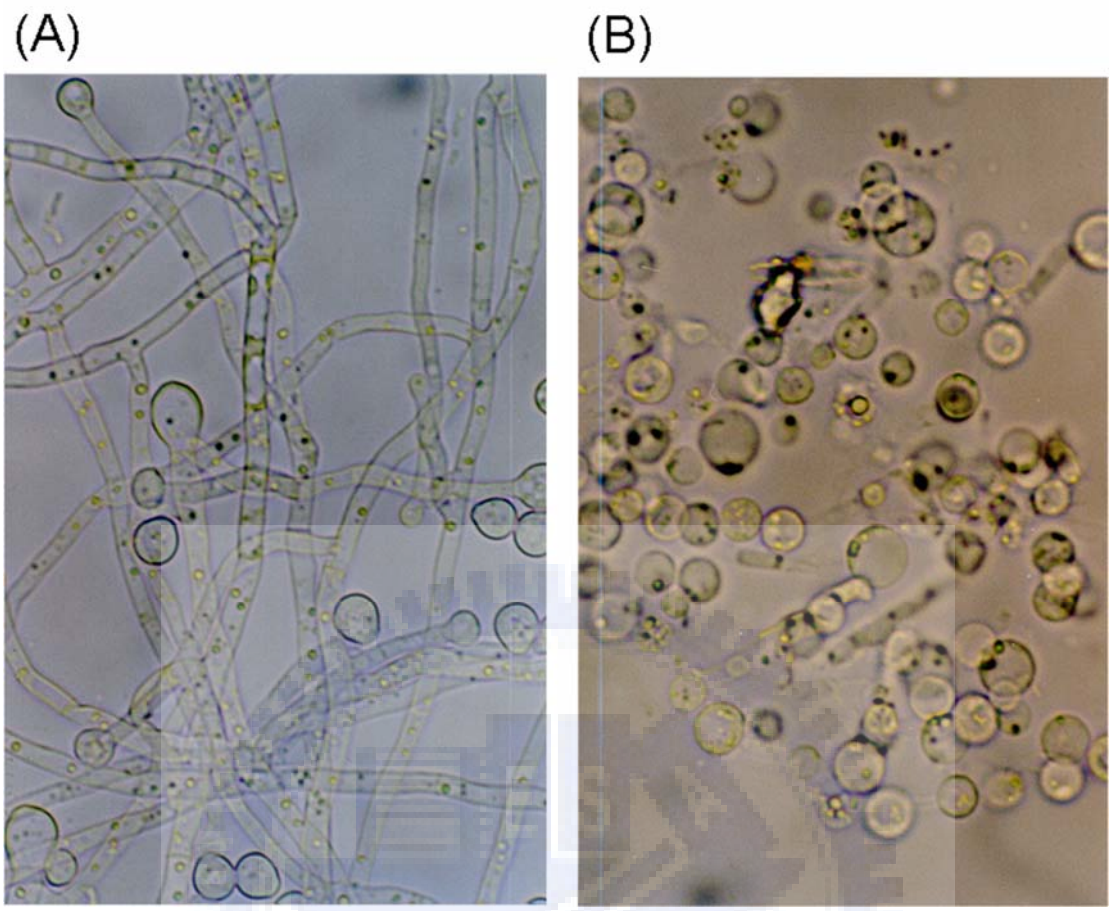


Figure 1. Morphology of *M. pilosus*. (A) Growth of wild-type *M. pilosus* in PDB medium was observed under microscope. (B) The mycelia of *M. pilosus* were digested by 2 mg Yatalase/ml, 2 mg lysing enzyme/ml and 2 μl β-glucuronidase/ml. They were converted into protoplasts after 4~5 hr digestion. The photographs were taken at a magnification of X1000.

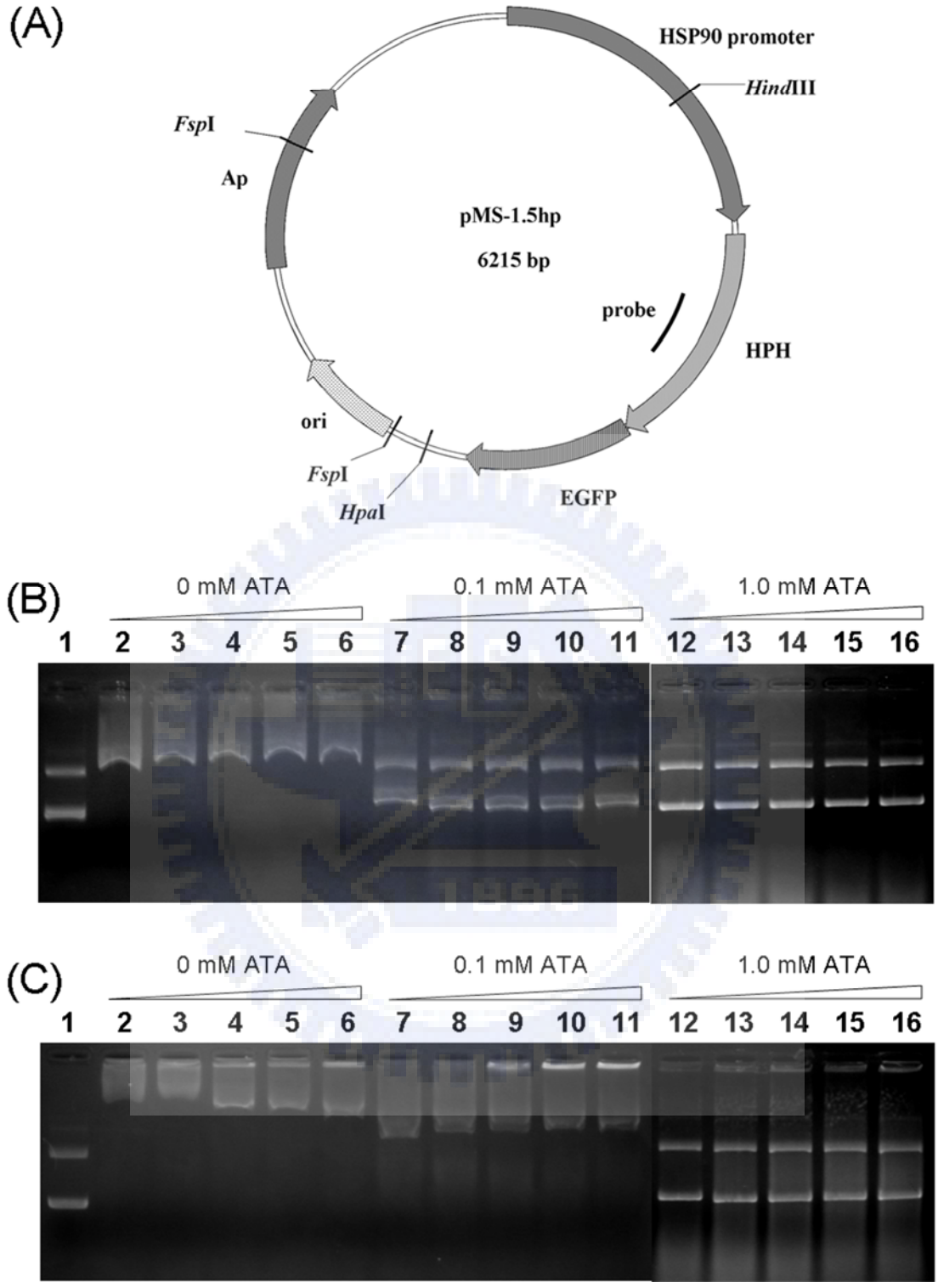


Figure 2. Map and Stability of pMS-1.5hp Plasmid. (A) Map of pMS-1.5hp plasmid used to overexpress a fusion protein of the hygromycin B resistance gene (HPH) with enhanced green fluorescent protein (EGFP). The pMS-1.5hp plasmid was modified

from the pHygEGFP plasmid. The small black bar represents the probe for analyzing integrations by Southern hybridization. A time course of the stability of the pMS-1.5hp plasmid in the *M. pilosus* BCRC38072 lysate was analyzed by including total protein of 1 mg ml^{-1} (B) or 3.5 mg ml^{-1} (C). Five μg of the pMS-1.5hp plasmid was incubated at 37^{o}C with the lysates for 5, 15, 30, 45, and 60 min, containing 0, 0.1, and 1.0 mm ATA. The plasmid DNA was separated on 1.0% agarose gel by electrophoresis. Lane 1, untreated control plasmid without lysate; lanes 2-6, plasmid DNA with lysate protein containing 0 mm ATA incubated for 5, 15, 30, 45, and 60 min; lanes 7-11, plasmid DNA with lysate protein containing 0.1 mm ATA incubated for 5, 15, 30, 45, and 60 min; lanes 12-16, plasmid DNA with lysate protein containing 1.0 mm ATA incubated for 5, 15, 30, 45, and 60 min.

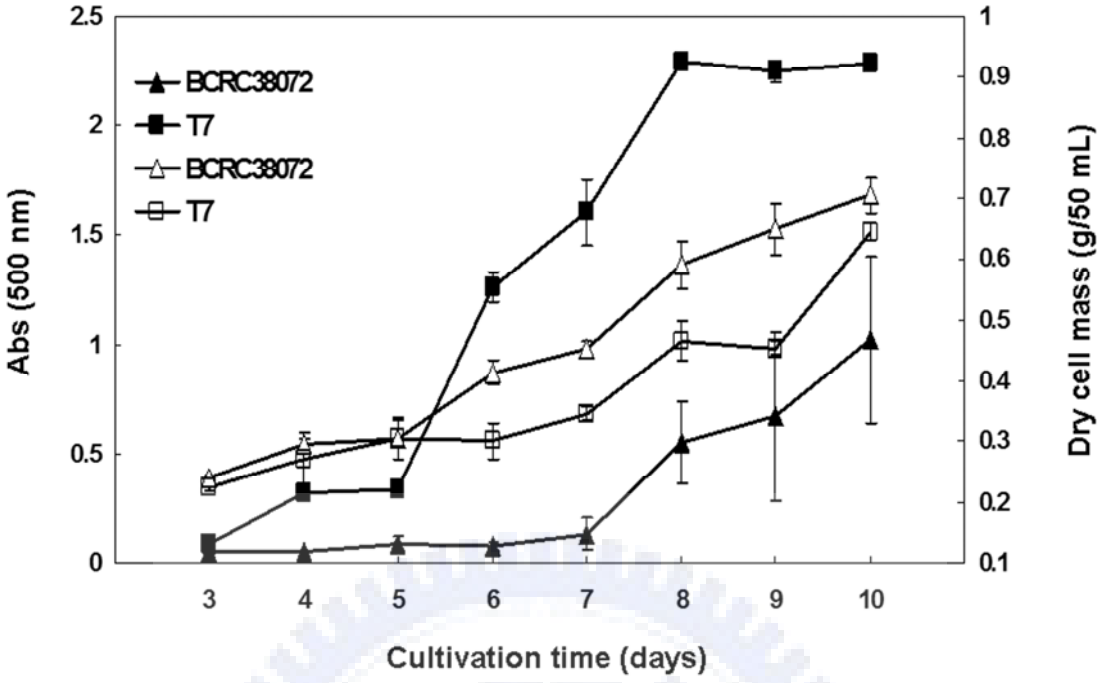


Figure 3. Submerged cultures of the wild type BCRC38072 and the T7 transformant were incubated on a 200 rpm rotary shaker for 10 days at 25 ^{o}C: (△) dry cell mass of the wild type; (□) dry cell mass of the transformant; (▲) red pigment production of the wild type; (■) red pigment production of the transformant. The red pigment level was detected by measuring the absorbance at 500 nm using a spectrophotometer.

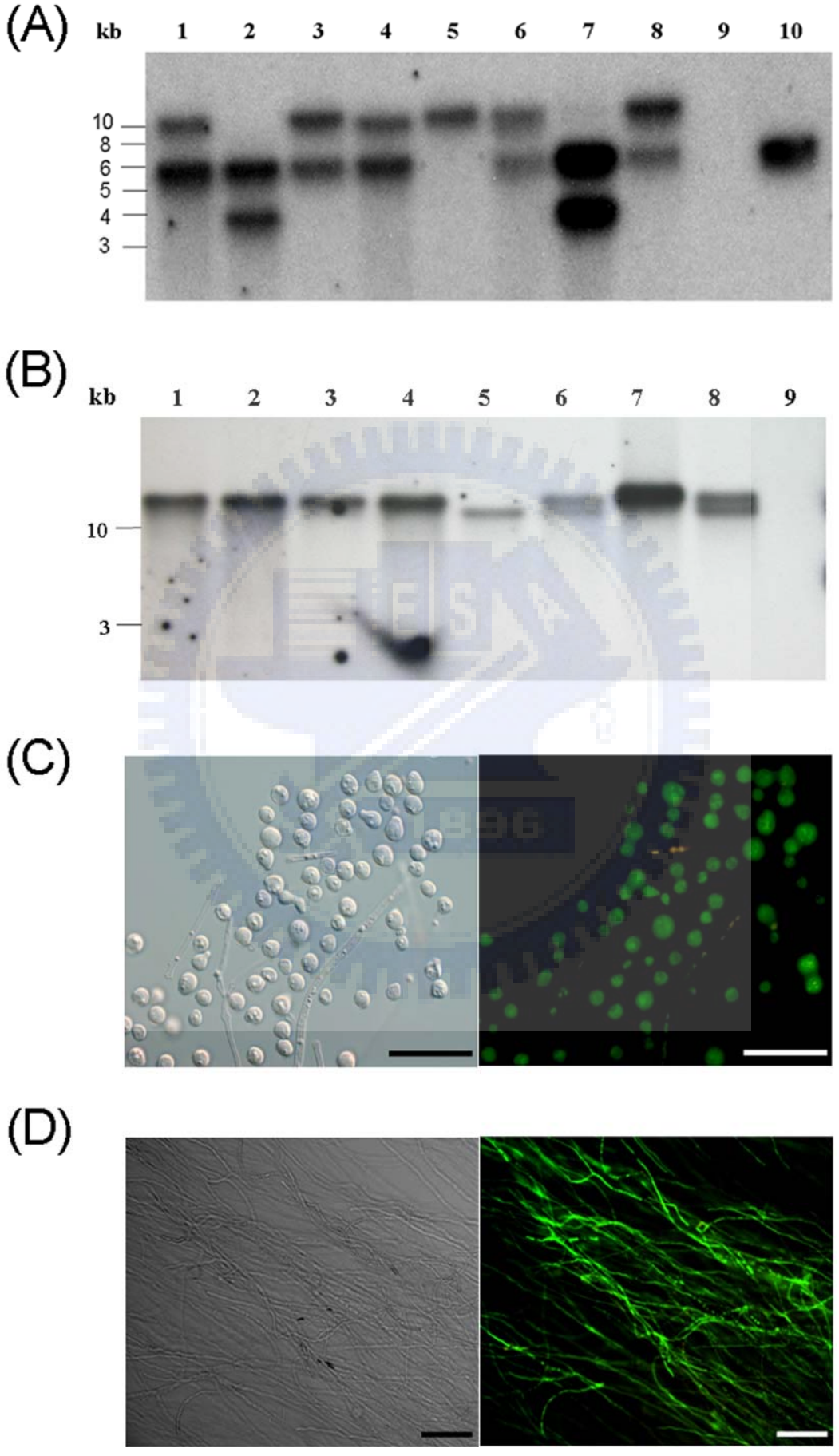

(A)
kb
1
2
3
4
5
6
7
8
9
10
10
8
6
5
4
3
(B)
kb
1
2
3
4
5
6
7
8
9
10
3
(C)
(D)

Figure 4. Analyses of Southern Hybridization and Fluorescence of *M. pilosus* transformants. Southern hybridization analyses of the genomes of eight transformants digested with *Hin*dIII (A) or *Pvu*II (B) and hybridized with an *hph* probe. Lanes 1-8 show DNA from the transformants (T1-T8); lane 9 shows DNA from wild-type *M. pilosus* BCRC38072; lane 10 shows the plasmid, pMS-1.5hp. Southern hybridization was performed using a DIG labeling system. Enhanced green fluorescent protein (EGFP) was expressed in the transformed *M. pilosus*. The conidia (C) and hyphae (D) of the transformant were analyzed by fluorescence (right panel) and bright field microscopy (left panel). The small bars represent distances of 50 μm and 100 μm in the conidia (C) and hyphae (D) respectively. The photograph was taken at a magnification of 1000X with an exposure time of 1 to 2 s.

Chapter 2

Cloning and Characterization of Monacolin K Biosynthetic Gene Cluster from *Monascus pilosus*

Abstract

Monacolin K is a secondary metabolite synthesized by polyketide synthases (PKS) from *Monascus*, and it has the same structure as lovastatin, which is mainly produced by *Aspergillus terreus*. In the present study, a Bacterial Artificial Chromosome (BAC) clone, mps01, was screened from the BAC library constructed from *Monascus pilosus* BCRC38072 genomic DNA. The putative monacolin K biosynthetic gene cluster was found within a 42-kb region in the mps01 clone. The deduced amino acid sequences encoded by the nine genes designated as *mokA* – *mokI*, which share over 54% similarity with the lovastatin biosynthetic gene cluster in *Aspergillus terreus*, were assumed to be involved in monacolin K biosynthesis. A gene disruption construct designed to replace the central part of *mokA*, a polyketide synthase gene, in wild type *M. pilosus* BCRC38072 with a hygromycin B resistance gene through homologous recombination resulted in a *mokA*-disrupted strain. The disruptant did not produce monacolin K, indicating that *mokA* encoded the PKS responsible for monacolin K biosynthesis in *M. pilosus* BCRC38072. Moreover, the transformant containing two copies of the *mokH* gene-encoded transcription factor showed higher production of monacolin K than wild type strain. Real-time RT-PCR analysis also demonstrated that the transcripts of monacolin K biosynthetic genes in the transformant were higher than those in wild type strain. These results suggested that *mokH* up-regulated the transcripts of monacolin K biosynthetic genes and the production of monacolin K.

Key words: Monacolin K, polyketide synthases, *Monascus pilosus*, Bacterial Artificial Chromosome

Introduction

Monascus spp. are filamentous fungi that have been used in Chinese fermented foods for thousands of years. They are known as producers of various secondary metabolites with polyketide structures, including monacolins, pigments, and citrinin (Endo, 1979; Endo et al., 1985; Endo et al., 1986; Shimizu et al., 2005). Monacolin K, a cholesterol synthesis inhibitor, was first isolated from the medium of *Monascus ruber* (Endo, 1979), and the same substance was found in *A. terreus* as lovastatin (Hendrickson et al., 1999). It belongs to polyketide synthesized by the iterative type I PKSs. The structure of monacolin K shares similarity with HMG-CoA; therefore, monacolin K competitively inhibits HMG-CoA reductase with HMG-CoA during cholesterol synthesis resulting in the reduction of cholesterol synthesis (Tobert, 2003).

In previous studies, the lovastatin biosynthetic pathway was proposed in *A. terreus*. Two polyketide synthases (*lovB* and *lovF*), transesterase (*lovD*), enoyl reductase (*lovC*), and P450 monooxygenase (*lovA*) have been proven to be involved in the structural biosynthesis of lovastatin (Hutchinson et al., 2000; Sorensen et al., 2003a; Sorensen et al., 2003b). Transformation of an extra copy of the *lovE* gene-encoded transcription factor into the wild type strain resulted in a 7- to 10-fold overproduction of lovastatin (Hutchinson et al., 2000). Although the lovastatin biosynthetic gene cluster in *A. terreus* has been characterized (Kennedy et al., 1999), the structural genes responsible for monacolin K (lovastatin) biosynthesis in *Monascus* are still unclear. In the present study, in order to explore the monacolin K biosynthetic gene cluster, construction of a BAC library from *M. pilosus* BCRC38072 producing monacolin K was carried out. According to the conserved region of the *lovB* gene (lovastatin nonaketide synthase, LNKS) in *A. terreus* was designed as a probe (Nicholson et al., 2001), the mps01 clone containing the putative monacolin K

biosynthetic gene cluster was isolated. Analysis of the disruption of the polyketide synthase gene (*mokA*) was conducted in order to identify the gene involved in monacolin K biosynthesis.

Materials and Methods

Strain used and growth conditions

M. pilosus BCRC38072, which is a monacolin K-producing strain isolated from red rice (anka), was collected from a local traditional market and used in this study. To identify the transcripts from monacolin K biosynthetic genes, the strain was incubated on YM (DIFCO 271120, Detroit, Michigan) agar for one week, and spore suspensions were obtained by washing cultured YM agar plates with distilled water. Mycelia were harvested after incubating 12 days at 25^{o}C with constant agitation in liquid medium (7% glycerol, 3% glucose, 3% monosodium glutamate, 1.2% polypetone, 0.2% $NaNO_3$ and 0.1% $MgSO_4.7H_2O$).

BAC library construction

The methods of Peterson et al. (2000) were used to construct the BAC library. Fragments of genomic DNA ranging in size from 150~300 kb were excised from pulse field gel electrophoresis (PFGE) and recovered by electroelution (BioRad, Hercules, CA). The eluted DNA was used for ligation. To perform the ligation, 100~200 ng of electroeluted DNA was mixed with 50 ng of linearized vector DNA (pIndigoBAC-5 *Hind*III Ready, Epicentre, Madison, WS), after which the ligation products were used to transform *Escherichia coli* strain TransforMax^{TM} EC 100^{TM} electrocompetent (Epicentre, Madison, WS) by electroporation. Transformed cells

were cultured on LB agar plates supplemented with chloramphenicol (12.5 μg mL^{-1}), IPTG (100 μg mL^{-1}) and XGal (50 μg mL^{-1}). The resulting white bacterial colonies were harvested and transferred to 384-well plates for library screening or storage at –80°C in freezing medium (2.5% [w/v] LB, 36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM sodium citrate, 0.4 mM $MgSO_4$, 6.8 mM $[NH_4]_2SO_4$ and 4.4% v/v glycerol).

Library screening, sequencing and sequence analysis

About 12,000 clones from the BAC library were cultured on LB agar plates, after which they were transferred to nylon membranes and then subjected to alkali-SDS lysis. The plasmid DNAs extracted from BAC clones were cross-linked to nylon membranes by UV irradiation. According to the conserved region of the ketosynthase domain of the *lovB* gene in *A. terreus* (Kennedy et al., 1999), the primer set (Mplov1: 5'-TCCACTGCCGTTTATGTTG-3'; Mplov2: 5'-GATGGGGTGAAGATGACGA-3') was designed for the probe synthesis using a PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The clones of membranes were screened to find a gene cluster involved in polyketide biosynthesis metabolism. Twenty-five positive BACs were identified in screening by colony hybridization. Two of these, mps01 and mps02, containing the longest inserted DNA were sequenced. BAC DNA for shotgun sequencing was extracted with a Qiagen Large-Construct kit (Qiagen, Valencia, CA), and sheared by sonication. The DNA fragments were blunted with the Bal31 nuclease and T4 DNA polymerase. Fragments were excised from the gel ranging in size from 1~2 kb and inserted into a pUC18/SmaI/CIAP (Amersham Pharmacia Biotech, Piscataway, NJ) vector. The ligation products were transformed into *E. coli* strain TransforMax™ EC 100™ electrocompent (Epicentre, Madison, WS) by electroporation to construct the subclone library. Transformed cells were cultured on LB agar plates supplemented

with ampicillin (50 μg mL^{-1}), IPTG (100 μg mL^{-1}), and XGal (50 μg mL^{-1}). The resulting white bacterial colonies were harvested and transferred to 96-deep well plates for overnight culture in LB broth and ampicillin (50 μg mL^{-1}). The inserts of these subclones were isolated and dissolved in 30 μl TE. Cycle sequencing reactions were carried out using a BigDye, V3.0 kit with universal primers. DNA sequencing for 10-fold coverage was performed with an ABI Prism 3700 Sequencer (Applied Biosystems, Foster City, CA). The Phred-Phrap-Consed system developed by the Phil Green Laboratory was used to assemble DNA fragments (Ewing and Green, 1998; Gordon et al., 1998). Nucleotide and deduced amino acid sequences were used to interrogate the non-redundant database at GenBank using BLASTN and BLASTX. The BAC mps01 contained the complete monacolin K gene cluster instead of the incomplete mps02.

Nucleic acid manipulations

Fungal genomic DNA was isolated by liquid nitrogen treatment according to the method developed by Bingle et al. (1999). Colony hybridization, Southern hybridization, and Northern hybridization were performed using the DIG system (DIG wash and buffer set) (Roche Diagnostics, Mannheim, Germany). The manipulations of transfer, immobilization and hybridization of DNA and RNA were carried out as described in Sambrook et al. (1989). Total RNAs of *M. pilosus* BCRC 38072 were isolated with TRIzol reagent (Invitrogen, Carlsbad, CA) for Northern hybridizations and reverse-transcription PCR analyses. The probes of monacolin K biosynthetic genes were DIG-labeled by PCR amplification using the PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The primer sets of monacolin K biosynthetic genes were pmkA-f: ATAGCTCCGAGAATGGTCCC, pmkA-r: CCATCAAGGATGCTCTGTC; pmkB-f: CTAGACTTTGCTTCCCACGCC, pmkB-r:

CATTGTCGAGCGTTGGAGTC; pmkC-f: TCAGAGATCTTCGTCCCGAC, pmkC-r: GGCCTGAGCCGAAGAAGTAC; pmkD-f: TGATGACTTTGCCCTGGCGG, pmkD-r: TCACCCAATGACTCTAGCCC; pmkE-f: TTCTCTCCCGACAACTGCCC, pmkE-r: AATGGTCACCGCCGACTGGA; pmkF-f: GCCCCGAATCCTACATGAAG, pmkF-r: GGCCCACCGTAGTTGATGTG; pmkG-f: CCTCGCTCTGAATATGACCC, pmkG-r: TCGGATCGGCTTCTCAAACC; pmkH-f: ACCTCATCGCTCCAGACCAT, pmkH-r: CTGCGAGAGAATGAGAGTGC; and pmkI-f: CCATACATTCTACCTTGCGG, pmkI-r: CTAGACTCGTTCATCGCGGC.

cDNA analysis

cDNA sequencing of nine genes, designated as *mokA-mokI,* was carried out to characterize their structure. First strand cDNA was synthesized by the ImProm-II™ Reverse Transcription System (Promega, Madison, WI) and used as the template for PCR. Amplification of full-length or partial cDNAs was performed with several sets of oligonucleotide primers. Sequences analyses were performed using VectorNTI 9.0 (InforMax, Frederick, MD) software.

Targeted gene disruption of *mokA*

The human cytomegalovirus (CMV) promoter of plasmid pHygEGFP (BD Biosciences Clontech, Palo Alto, CA) was replaced with a 0.5-kb *Bgl*II-*Xho*I fragment of the heat shock protein 90 (*hsp 90*) promoter from *M. pilosus* BCRC38072 (GenBank accession no. DQ983312) to obtain the plasmid pMS-hsp. The *hph* cassette, a hygromycin B resistance gene, was flanked at the 5' site by 2.0 kb and at the 3' site by 3.1 kb, respectively, of the *mokA* gene to obtain the plasmid pMkAko. The transformant that had *mokA* replaced with pMkAko by homologous recombination was identified by Southern hybridization.

Transformation of *M. pilosus* BCRC38072

The conidia from a one-week culture of *M. pilosus* were incubated in 100 ml of Vogel medium at 30 ^{o}C for 16~18 h. The mycelia were harvested on miracloth (Millipore, Bedford, USA) and washed in MA digestion solution (0.1 M maleic acid, pH 5.5 and 1.2 M $(NH_4)_2SO_4$). The mycelia were digested for 4~5 h using 100 mg Yatalase (Takara, Tokyo, Japan), 100 mg lysing enzyme (Sigma, St. Louis, USA) and 100 μl β-glucuronidase (Sigma, St. Louis, USA) in 50 ml MA digestion solution. To remove undigested mycelia, protoplasts were harvested by passing through miracloth and by centrifugation at 1000 rpm for 10 min (Sorvall, Wilmington, USA). The protoplasts were maintained in 80 % STC (1 M sorbitol, 50 mM Tris pH 8.0, 50 mM $CaCl_2$) and 20 % PTC (40 % PEG 4000, 50 mM Tris pH 8.0, 50 mM $CaCl_2$), and dimethylsulfoxide (DMSO) was added to a final concentration of 1 %. For genetic transformation of *M. pilosus*, the 100 μl of protoplasts were mixed with 5 μg of linearized DNA. The mixtures were incubated on ice for 30 min. One milliliter of PTC was added and mixed gently. Following incubation at room temperature for 20 min, the protoplast mixtures were added to 15 ml of SYP medium (1 M sorbitol, 0.1 % yeast extract, 0.1 % peptone and 2 % agar) that contained 60 μg hygromycin B/ml. Plates were incubated at 28^{o}C and the transformants were detected by PCR and Southern hybridization.

Construction of *mokH* gene expression vector

The glyceraldehyde-3-phosphate dehydrogenase (*gpd*) promoter of plasmid pAN7-1 (Punt et al., 1987) from *A. nidulans* was replaced with a 1.1-kb *Bgl*II-*Sal*I fragment of the *gpd* promoter from *M. pilosus* BCRC38072 (GenBank accession no. DQ984142) to obtain the plasmid pMS. The PCR product of 1.5-kb *mokH* genomic

DNA was introduced into the plasmid pMS with digestion of the *Sal*I-*Bam*HI site blunted by End-It DNA End-repair kit (Epicentre, Madison, WS) to obtain the plasmid pMS-mkH. The PCR product of a 3.5-kb fragment containing the promoter, *mokH* gene and terminator from pMS-mkH was introduced into the plasmid pMS-hsp with digestion of the *Cla*I site blunted by End-It DNA End-repair kit (Epicentre, Madison, WS) to obtain the plasmid pMSmokH.

Quantitation of monacolin K biosynthetic gene expression by real-time RT-PCR

Mycelia of *M. pilosus* were harvested after incubating for 8 days at 25^{o}C and ground to a fine powder in liquid nitrogen with a mortar and pestle. Total RNA from 0.2 g of mycelial powder was prepared by TRIzol reagent (Invitrogen, Carlsbad, CA) and treated with 2 U of RNase-free DNase for 1 h. First strand cDNA was synthesized by ImProm-II^{TM} Reverse Transcription System with random hexamers (Promega, Madison, WI) at 42^{o}C for 1 h and used as the template for real-time PCR. Real-time PCR was carried out using the Absolute SYBR Green ROX Mix (Abgene, Epsom, UK) in a final volume of 25 μl for each reaction in the ABI Prism 7000 sequence detection system (Applied Biosystems, Foster City, CA). The thermal cycling parameters consisted of an initial heating at 95^{o}C for 15 min, denaturing at 95^{o}C for 10 s, and annealing and extension at 60^{o}C for 1 min, with an amplification of 40 cycles. The primer set of *M. pilosus* 18S rRNA designed by the PrimerExpress software was used as a positive control for normalization. The primer set of 18S rRNA was mpf18SrRNA: TCTCGTAATCGGAATGAGAACGA, and mpr18SrRNA: TACGCTATTGGAGCTGGAATTACC. RNA samples without a reverse transcriptase step were also analyzed to determine genomic DNA contamination. The primer sets of monacolin K biosynthetic genes were mkA-f/QPCR: CTGGTGCAGACACAGTACGACAT, mkA-r/QPCR:

GGAACCATCGCCGACAAAT; mkB-f/QPCR: GGGACCCTGAGTTTCGAACA, mkB-r/QPCR: GCACTTTTTCACCCCGTTGA; mkC-f/QPCR: GAGGCCAGCGCGACAAT, mkC-r/QPCR: GTGACAGTGCGTGTCACCAAA; mkD-f/QPCR: AGCAGCAGATGGGTAGGATACC, mkD-r/QPCR: ACCTCCTCCCGCATATACTGAGT; mkE-f/QPCR: GCGACGATTGTGATGCAGAT, mkE-r/QPCR: ATCTTCTGCGCCGTGCTTT; mkF-f/QPCR: AACGGAGAAGCAGATGAACCA, mkF-r/QPCR: TCCCACCAAGCCCAAAACT; mkG-f/QPCR: CGTCCGGAAGGTCCTGAAG, mkG-r/QPCR: TGAACCCCCCCATACTACCA; mkH-f/QPCR: GGAGTGGCCAAAACAGGAAA, mkH-r/QPCR: TGCGGGTGTTGGATTGTTG; and mkI-f/QPCR: TGCTGGGAGGTGCTTTTACC, mkI-r/QPCR: AATGTGGATGGCGAGAAGGA. Measurements of gene expression were performed three times and the mean of these values was used for analysis. The relative gene expression level was calculated by the $2^{-\Delta\Delta CT}$ method as described in Applied Biosystems User Bulletin No. 2.

Measurement of monacolin K

The aliquots of *M. pilosus* culture were cleared of cells and filtered through a 0.2 mm filter. The supernatants were analyzed by high performance liquid chromatography (HPLC) performed on a Waters system (Waters, Milford, MA) fitted with a reverse-phase C_{18} column (LichroCART 250-4, Rp-18e, 5 μm). The HPLC parameters were as follows: solvent A, 0.1% phosphorus acid in water; solvent B, acetonitrile; 35 % A and 65 % B in 30 min; flow rate, 1.5 mL min^{-1}; and detection by UV spectroscopy (Waters 600 pump and 996 pgotodiode array detector). The monacolin K purified from the cultivation of *M. pilosus* BCRC38072 was verified by mass and ^{1}H NMR spectroscopic analyses. A standard monacolin K compound

(Sigma, St. Louis, USA) was used to confirm the analysis by HPLC.

Nucleotide sequence accession number

The nucleotide sequence of the monacolin K biosynthetic gene cluster has been submitted to GenBank under the accession number DQ176595.

Results

Cloning of monacolin K biosynthetic gene cluster

Studies on fungal polyketide biosynthetic genes indicate that metabolites are largely synthesized by iterative multifunctional polyketide synthase systems (Pfeifer and Khosla, 2001). Each PKS minimally carries keto-synthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains to catalyze different modifications. To search for the genes related to monacolin K biosynthesis, degenerate primers designed according to the conserved region of the KS domain of the *lovB* gene in *A. terreus* (Nicholson et al., 2001) were used to amplify genomic DNA from *M. pilosus* BCRC38072. The candidate PCR products were sequenced and analyzed. The result of the amplified DNA showed that the PCR product shared a 75% similarity with the KS domain of the *lovB* gene in *A. terreus* (data not shown). This DNA fragment was further used to design a specific probe for cloning of the PKS gene.

A BAC library consisting of 12,000 clones was constructed from the total DNA of *M. pilosus* BCRC38072. By screening the library with the specific PKS probe, 25 positive BACs were identified. To evaluate the sizes of the BACs, pulsed field electrophoresis gel (PFEG) and Southern hybridization were carried out. The BAC designated as mps01 was selected for shotgun sequencing. It yielded a contig of

approximately 160 kb. Database searches and open reading frame prediction further provided information on the putative gene loci. The whole sequences of mps01 were annotated by BLASTN and BLASTX and 30 open reading frames (ORFs) were predicted.

Identification of genes involved in monacolin K biosynthesis

Within the 30 putative genes, nine genes were found to have strong homology to the genes involved in the lovastatin biosynthetic gene cluster (Kennedy et al., 1999), and they were presumed to encode proteins required for monacolin K biosynthesis (**Table 1**). The putative gene cluster of monacolin K covered 42 kb (**Figure 1A**) (GenBank accession no. DQ176595). Moreover, they also shared high similarity with the genes involved in the compactin biosynthetic gene cluster of *Penicillium citrinum* (Abe et al., 2002). The structure of monacolin K differs from that of compactin, in which a methyl group derived from S-adenosyl-L-methionine (SAM) is introduced at the C-6 position of the nonaketide-derived backbone. An extensive comparison analysis of these nine genes indicated the presence of two polyketide synthase genes by BLAST. One was predicted to be responsible for the synthesis of the nonaketide skeleton (*mokA*), while the other was for the synthesis of the diketide skeleton (*mokB*). Also included were a P450 monooxygenase gene (*mokC*), an oxidoreductase gene (*mokD*), a dehydrogenase gene (*mokE*), a transesterase gene (*mokF*), an HMG-CoA reductase gene (*mokG*), a transcription factor gene (*mokH*), and an efflux pump gene (*mokI*).

To assess the transcription of the predicted monacolin K biosynthetic genes, Northern hybridizations were performed with DIG-labeled probes. Transcripts of the putative monacolin K biosynthetic genes could be detected on the eighth day (**Figure 1B**). The cDNA sequences were used to annotate the nine gene sequences.

For this, several sets of oligonucleotide primers were designed to amplify cDNAs by reverse-transcription PCR. Only the *mokD* gene revealed no intron, and others contained at least one intron with sizes ranging from 52 bp to 109 bp. The deduced amino acid sequences of the putative monacolin K biosynthetic gene cluster were confirmed, and the sequence similarity among corresponding genes of *A. terreus* and *P. citrinum* is shown in **Table 1**. Several conserved domains were recognized in MokA and MokB by comparing their amino acid sequences with those of known PKSs. The domains of keto-synthase (KS), acyltransferase (AT), dehydratase (DH), methyltransferase (MeT), keto-reductase (KR), and acyl carrier protein (ACP) are included in both MokA and MokB. Additionally, MokB comprised an additional enoyl reductase domain (ER) similar to the corresponding gene *lovF* of *A. terreus*, as shown in **Figure 1C**. The *mokH* gene-encoded transcription factor was suggested to be a positive regulatory protein for the production of monacolin K just like *lovE* which is involved in lovastatin biosynthesis in *A. terreus* (**Figure 2**) (Kennedy et al., 1999). The arrangement of a cysteine-rich nucleotide-binding domain indicated that the consensus sequence $CX_2CX_6CX_{11}CX_2CX_6C$ represented a Zn_2Cys_6 type zinc finger (Hutchinson et al., 2000; Kennedy et al., 1999).

Polyketides are frequently synthesized from CoA thioesterified carboxylic acids and the extent of keto-group processing varies from one condensation cycle to another (Hutchinson et al., 2000). In this study, the phylogeny was further constructed according to the conserved domain of keto-synthase (**Figure 3A**). The result showed that the PKSs were divided into three clades. The clade of *mokA* and *mokB* was subdivided into two subclades belonging to the structural type of highly reduced polyketide. Since the MeT domain in *mlcA* of the compactin biosynthetic gene cluster is assumed to be inactive (Abe et al., 2002), the domain was compared among corresponding PKSs from *M. pilosus*, *P. citrinum* and *A. terreus*, and our results

showed that the amino acid residues of *mlcA* in consensus motifs were different from those of the other PKS, as boxed (**Figure 3B**).

Disruption of *mokA* gene in *M. pilosus* BCRC38072

The *mokA* gene-encoded polyketide synthase was suggested to synthesize the nonaketide of monacolin K. In this study, the *mokA* gene was disrupted in *M. pilosus* BCRC38072 by homologous recombination to identify the gene involved in monacolin K biosynthesis. Plasmid pMkAko was linearized at the *Fsp*I sites and was used to transform strain BCRC38072 (**Figure 4A**). Forty-four transformants were isolated, and one of which had a MokA$^-$ genotype by Southern hybridization.

Southern hybridization analysis of *Nde*I-digested DNA indicated that instead of a 7.1-kb fragment corresponding to the *mokA* gene in wild type BCRC38072, a 3.8-kb *Nde*I fragment was present in the disruptant BCRC38135 (**Figure 4BC**). Thus, our result revealed that disruption of *mokA* gene had occurred. A precise gene replacement, in which the nonfunctional *mokA* gene construct (pMkAko) replaced the functional chromosomal *mokA* gene, yielded *Nde*I fragments of 3.8 kb. In addition, the amount and structure of monacolin K produced from *M. pilosus* BCRC38072 and BCRC38135 on the eighth day of cultivation were further determined by HPLC, mass, and ^{1}H NMR spectroscopic analyses. Monacolin K was detectable in wild type *M. pilosus* BCRC38072, as confirmed by UV light absorption, mass, and ^{1}H NMR spectra (**Figure 5**). The peak of monacolin K from *M. pilosus* BCRC38072 was identified by comparison to the monacolin K standard, which showed three maximum absorptions at (λ_{max}) 230, 237 and 246 nm (**Figure 5A**). The mass spectrum of monacolin K revealed that the molecular weight was 404, which also agreed with the standard ($C_{24}H_{36}O_5$) (**Figure 5B**). Moreover, the structure of monacolin K was further verified by ^{1}H NMR spectrum [^{1}H NMR (400MHz, $CDCl_3$): δ 5.98 (1H, d, *J* =

10.0 Hz, H-5), 5.76 (1H, dd, *J* = 6.4, 6.0 Hz, H-6), 5.50 (1H, d, *J* = 2.8 Hz, H-4), 5.36 (1H, dt, *J* = 4.8, 3.2 Hz, H-1), 4.60 (1H, m, H-5'), 4.33 (1H, m, H-3'), 2.72 (1H, dd, *J* = 5.2, 4.8 Hz, H_{ax} -2'), 2.62 (1H, ddd, *J* = 3.6, 3.2, 2.4, H_{eq} -2'), 2.43 (2H, m, H-3), 2.38 (1H, m, H-7), 2.36 (1H, m, H-2''), 2.27 (1H, dd, *J* = 2.8, 2.8 Hz, H-8a), 1.98 (1H, m, H_{eq} -4'), 1.95 (2H, m, H-2), 1.89 (1H, m, H-6'), 1.72 (1H, m, H-8), 1.65 (1H, m, H_{ex} -4'), 1.63 (1H, m, H-3''), 1.48 (1H, m, H-7'), 1.42 (1H, m, H-3''), 1.38 (1H, m, H-7'), 1.29 (1H, m, H-6'), 1.11 (3H, d, *J* = 7.2 Hz, H-2''-CH_3), 1.06 (3H, d, *J* = 7.2 Hz, H-3-CH_3), 0.88 (3H, d, *J* = 7.2 Hz, H-7-CH_3), 0.86 (3H, t, *J* = 7.6 Hz, H-4'')]. However, the disruptant *M. pilosus* BCRC38135 did not produce monacolin K, indicating that *mokA* gene is responsible for monacolin K biosynthesis in *M. pilosus* BCRC38072 (**Figure 5C**).

Overexpression of *mokH* gene in *M. pilosus* BCRC38072

The *mokH* gene-encoded transcription factor, a Zn_2Cys_6 binuclear protein, was suggested to be a positive regulatory protein for the production of monacolin K just like *lovE* which is involved in lovastatin biosynthesis in *A. terreus* (Hutchinson et al., 2000; Kennedy et al., 1999). In this study, three transformants of *M. pilosus* BCRC38072 were obtained through the transformation of the plasmid pMSmokH containing the *mokH* gene (**Figure 6**). All of these transformants showed higher monacolin K production than the wild type. Three transformants possessed the complete *gpd* promoter, *mokH* gene, and *trpC* terminator which were verified by PCR, nest-PCR and sequencing. The transcripts of *mokH* gene of these transformants were higher expression than those of the wild-type by Northern blot analysis (**Figure 7**).

The copy numbers of the *mokH* gene in wild type (BCRC38072) and three transformants (T-mokH1, T-mokH2 and T-mokH3) were determined by Southern

hybridizations using restriction enzymes *Sal*I, *Xho*I, *Hin*dII and *Nar*I (**Figure 8A**). *Sal*I and *Xho*I cut at one site in the pMSmokH plasmid, while *Hin*dII and *Nar*I did not cut within the plasmid. The 3.2-kb *Sal*I and 2.5-kb *Xho*I fragments corresponded to the single copy of *mokH* in wild type BCRC38072. Moreover, besides the fragments corresponding to the wild-type, the transformants revealed two integration events (two copies) using *Sal*I and *Xho*I restriction enztmes. In addition, the 3.4-kb *Hin*dIII and 5-kb *Nar*I fragments corresponding to the single copy of *mokH* in wild type BCRC38072 were found. The transformants, T-mokH1 and T-mokH3, also showed two integration events when *Hin*dIII and *Nar*I were used as the restriction enzymes. However, only one fragment corresponding to transformant T-mokH2 was detected. The results suggest that the plasmid (pMSmokH) integrated as a tandem repeat in T-mokH2 (**Figure 8B**), while it randomly integrated into ectopic locations of chromosome in the transformants T-mokH1 and T-mokH3. These results indicated that three transformants contained two copies of *mokH*. The concentration of monacolin K produced from the transformant T-mokH1 on the time course of 15 days cultivation was higher than that of wild-type. The transformant revealed 1.7-fold higher monacolin K productivity than the wild type BCRC38072, while the cultured cell mass revealed no significant increase (**Figure 9**). These results suggested that *mokH* up-regulated the production of monacolin K.

Quantitation of monacolin K biosynthetic gene expression by real-time RT-PCR

To access monacolin K biosynthetic gene expression in wild type (BCRC38072) and the transformant (T-mokH1), real-time RT-PCR experiments were conducted. The data from each sample treatment were normalized against those for the 18S rRNA positive control. The results of real-time RT-PCR for monacolin K biosynthetic genes are shown in **Table 2**. The overexpression of *mokH* gene in the transformant

T-mokH1 resulted in a 1.2- to 3.0-fold increase in the transcript levels of the monacolin K biosynthetic genes. These results were generally in agreement with the monacolin K productivities of the transformants described above. It suggested that *mokH* up-regulated the transcripts of monacolin K biosynthetic genes.

Discussion

Monacolin K, also known as lovastatin, is a polyketide used to reduce serum cholesterol levels in humans. Over the past years, it has become clear that polyketides are assembled in a variety of mechanistically complex ways (Hutchinson et al., 2000). Studies on the lovastatin biosynthetic gene cluster of *A. terreus* have shown 18 putative open reading frames based on the sequence alignment and characterization of genetically related fungal strains (Kennedy et al., 1999). Surprisingly, only nine genes in the BAC of *M. pilosus* BCRC38072 have revealed high homology to the genes involved in lovastatin biosynthetic gene cluster of *A. terreus* (**Table 1**). Moreover, the genomic arrangement of monacolin K biosynthetic genes in *M. pilosus* BCRC38072 has corresponded to the lovastatin biosynthetic genes in *A. terreus* (**Figure 1A**). The high homology between gene clusters of *mok* and *lov* implies that *mok* gene cluster was responsible for monacolin K biosynthesis. To prove this, we disrupted the *mokA* gene-encoded polyketide synthase from wild type *M. pilosus* BCRC38072. The phenotype of lost monacolin K productivity in the disruptant BCRC 38135 indicates that the *mokA* gene was essential for monacolin K production (**Figure 5C**).

In particular, these genes also showed significant homology to genes identified in the compactin biosynthetic gene cluster of *Penicillium citrinum* (Abe et al., 2002a).

However, the genomic arrangement of compactin biosynthetic genes was different from that of the monacolin K or lovastatin biosynthetic gene clusters (**Figure 1A**). The order and direction of P450 monooxygenase (*mokC)*, polyketide synthase (*mokA*), oxidoreductase (*mokD*), dehydrogenase (*mokE*), and transesterase (*mokF*) was the same in *M. pilosus*, *A. terreus* and *P. citrinum*, whereas the organization of other genes of the compactin biosynthetic gene cluster was different (Kennedy et al., 1999; Abe et al., 2002a). Furthermore, polyketide synthase (*mokB*), monooxygenase (*mokC*), oxidoreductase (*mokD*), dehydrogenase (*mokE*), and an efflux pump (*mokI*) appeared to have the same number of introns and similar intron positions among *M. pilosus*, *A. terreus* and *P. citrinum*.

A. terreus and *P. citrinum* both belong to the family Trichocomaceae, but they are different from *M. pilosus* which belongs to the family Monascaceae (Kuraishi et al., 2000). Interestingly, lovastatin biosynthetic genes from *A. terreus* revealed a higher homology with monacolin K biosynthetic genes from *M. pilosus* than with compactin biosynthetic genes from *P. citrinum*. Since polyketides play an ecological role in the environment regarding microbial competition, genetic differences might reflect extreme environmental stress and subsequent genetic changes in these species (Duffy et al., 2004). In addition, many of the predicted PKSs in the PKS clade that produce highly reduced polyketides have divergent and presumably nonfunctional methyltransferase domains (Kroken et al., 2003). The structure of monacolin K differs from that of compactin, in which a methyl group derived from S-adenosyl-L-methionine (SAM) is introduced at the C-6 position of the nonaketide-derived backbone (Abe et al., 2002a). The MeT domain in *mokA* and *lovB* genes is assumed to be active instead of the *mlcA* gene. The consensus motif of MeT domain of *mlcA* were found different from the relative PKSs (*mokA* and *lovB*) in some amino acid residues, A→L, G→I, QM→HL, and I→T (**Figure 3B**).

Furthermore, there were two more introns located at the methyltransferase domain of *mlcA*, whereas *mokA* and *lovB* contained the same number of introns and similar introns positions. Therefore, the differences among amino acid residues could be the reason for the lack of methyltransferase activity of *mlcA* in *P. citrinum*. These results could form the basis for the study of site-directed mutagenesis in order to understand the methyltransferase activity of PKSs (Abe et al., 2002a). Among these genes shown in **Table 1**, the transcription factor (*mokH*, *lovE* and *mlcR*) and HMG-CoA reductase (*mokG*, *lvrA* and *mlcD*) were found lesser similarities to each other. The number and positions of introns were also different from one another. Nevertheless, the transcription factor and HMG-CoA reductase genes were assumed to be regulators responsible for up-regulation and down-regulation, respectively (Hutchinson et al., 2000; Abe et al., 2002b). HMG-CoA reductase (*mokG*) could play a role in conferring resistance to monacolin K (Abe et al., 2002b), and theoretically there was no effect upon the structure of the polyketides (Hutchinson et al., 2000).

The regulation of secondary metabolite clustered genes is complex. According to the proposed model of the signal transduction pathway that regulates the biosynthesis of polyketide Sterigmatocystin (ST, the penultimate precursor to aflatoxin B1) in *A. nidulans*, G-protein and transcription factor regulation are also involved in fungal secondary metabolism (Shimizu et al., 2003; Yu et al., 2004). In this study, *mokH* gene shared 54% similarity with *lovE* gene-encoded transcription factor from *A. terreus* (**Table 1**). Three transformants containing two copies of the *mokH* gene-encoded transcription factor were obtained, and this resulted in the improvement of monacolin K production. The transcripts of monacolin K biosynthetic genes were also found to increase (**Table 2**). These results were consistent with the finding of Kennedy et al. (Hutchinson et al., 2000; Kennedy et al.,

1999), who demonstrated that the extra copy of the *lovE* gene improves the yield of lovastatin. The transcription factors, *mokH*, *lovE*, and *mlcR* genes, are Zn_2Cys_6 binuclear cluster proteins that possesses DNA-binding and activation domains typical of the GAL4-type family of positive regulatory proteins (Marzluf, 1997). The GAL4-type transcription factor binds to a palindromic consensus sequence such as 5'-CGGNxCCG-3' or 5'-CCGNxCGG-3', where Nx represents variable spacing. Different from the nucleotide-binding sequence of GAL4, AFLR which is the transcription factor of aflatoxin, a Zn_2Cys_6 binuclear protein, binds to the partially palindromic consensus sequence 5'-TCGN$_5$CGR-3' (Ehrlich et al., 2003). Here, within the region of the monacolin K biosynthetic gene cluster in *M. pilosus* BCRC38072, several consensus sequences 5'-TCGN$_5$CGR-3' were also found at the promoter region of different genes (GenBank accession no. DQ176595). Thus, it is possible that the expressions of monacolin K biosynthetic genes were influenced by the regulatory gene, *mokH*, with the same nucleotide-binding sequence. In addition, we demonstrated that the *mokH* gene was required for increasing monacolin K production. Further study on the binding sequence of the transcription factor such as electrophoretic mobility shift assay (EMSA) will clarify the individual regulation of each gene in monacolin K biosynthesis.

The data for the monacolin K biosynthetic gene cluster can provide important information about the biosynthesis of monacolin K (lovastatin) between *Monascus* and *Aspergillus*. It is interesting that polyketide synthases between *mok* genes and *lov* genes are othologs and also related in compactin biosynthesis by *Penicillium citrinum*. Thus, there are three othologous gene clusters in different fungi which are useful to study evolution of genes for secondary metabolism. Moreover, this suggests that the structural variety of polyketides produced by fungi accompanies the enzymatic variation (Hutchinson et al., 2000; Nicholson et al., 2001). Studies on the

regulation of fungal secondary metabolism and the development of novel polyketides have great potential for screening effective medications.

References

(1) Abe, Y.; Suzuki, T.; Ono, C.; Iwamoto, K.; Hosobuchi, M.; Yoshikawa, H. Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in *Penicillium citrinum*. *Mol. Genet. Genomics.* **2002a**, *267*, 636-646.

(2) Abe, Y.; Suzuki, T.; Mizuno, T.; Ono, C.; Iwamoto, K.; Hosobuchi, M.; Yoshikawa, H.; Effect of increased dosage of the ML-236B (compactin) biosynthetic gene cluster on ML-236B production in *Penicillium citrinum*. *Mol. Genet. Genomics.* **2002b**, *268*, 130-137.

(3) Bingle, L. E. H.; Simpson, T. J.; Lazarus, C. M. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet. Biol.* **1999**, *26*, 209-223.

(4) Duffy, B.; Keel, C.; Défago, G. Potential role of pathogen signaling in multitrophic plant-microbe interactions involved in disease protection. *Appl. Environ. Microbiol.* **2004**, *70*, 1836-1842.

(5) Ehrlich, K.C.; Montalbano, B.G.; Cotty, P.J. Sequence comparison of *aflR* from different *Aspergillus* species provides evidence for variability in regulation of aflatoxin production. *Fungal Genet. Biol.* **2003**, *38*, 63-74.

(6) Endo, A. Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* (Tokyo) **1979**, *32*, 852-854.

(7) Endo, A.; Hasumi, K.; Nakamura, T.; Kunishima, M.; Masuda, M. Dihydromonacolin L and monacolin X, new metabolites which inhibit cholesterol biosynthesis. *J. Antibiot.* (Tokyo) **1985**, *38*, 321-327.

(8) Endo, A.; Komagata, D.; Shimada, H. Monacolin M, a new inhibitor of cholesterol biosynthesis. *J. Antibiot.* (Tokyo) **1986**, *39*, 1670-1673.

(9) Ewing, B; Green, P. Base-calling of automated sequencer traces using phred. II.

Error probabilities. *Genome. Res.* **1998**, 8, 186-194.

(10) Gordon, D; Abajian, C; Green, P. Consed: a graphical tool for sequence finishing. *Genome. Res.* **1998**, 8, 195-202.

(11) Hendrickson, L.; Davis, C. R.; Roach, C.; Nguyen, D. K.; Aldrich, T.; McAda, P. C.; Reeves, C. D. Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem. Biol.* **1999**, *6*, 429-439.

(12) Hutchinson, C. R.; Kennedy, J.; Park, C.; Kendrew, S.; Auclair, K.; Vederas, J. Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases. *Antonie Van Leeuwenhoek.* **2000**, *78*, 287-295.

(13) Kagan, R. M.; Clarke, S. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* **1994**, *310*, 417-427.

(14) Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science.* **1999**, *284*, 1368-72.

(15) Kroken, S.; Glass, N. L.; Taylor, J. W.; Yoder, O. C.; Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. USA.* **2003**, *100*, 15670-15675.

(16) Kuraishi, H.; Itoh, M.; Katayama, Y.; Hasegawa, A.; Sugiyama, J. Ubiquinone systems in fungi. V. Distribution and taxonomic implications of ubiquinones in Eurotiales, Onygenales and the related plectomycete genera, except for *Aspergillus*, *Paecilomyces*, *Penicillium*, and their related teleomorphs. *Antonie Van Leeuwenhoek.* **2000**, *77*, 179-186.

(17) Marzluf, G.A. Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 17-32.

(18) Nicholson, T. P.; Rudd, B. A.; Dawson, M.; Lazarus, C. M.; Simpson, T. J.; Cox, R. J. Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem. Biol.* **2001**, *8*, 157-78.

(19) Peterson, D. G.; Tomkins, J. P.; Frisch, D. A.; Wing, R. A.; Paterson, A. H. Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. J. Agric. Genomics. 2000; Volume 5.

(20) Pfeifer, B. A.; Khosla, C. Biosynthesis of polyketides in heterologous hosts. Microbiol. *Mol. Biol. Rev.* **2001**, *65*, 106-118.

(21) Punt, P.J.; Oliver, R.P.; Dingemanse, M.A.; Pouwels, P.H.; van den Hondel, C.A. Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene.* **1987**, *56*, 117-124.

(22) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406-425.

(23) Sambrook, J., Fritsch, E., Maniatis, T. Molecular cloning: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, New York. **1989**. *1*, 7.39.

(24) Shimizu, K.; Hicks, J.K.; Huang, T.-P.; Keller, N.P. Pka, Ras and RGS protein interactions regulate activity of AflR, a Zn(II)2Cys6 transcription factor in *Aspergillus nidulans*. *Genetics.* **2003**, *165***,** 1095-1104.

(25) Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T. Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2005**, *71*, 3453-3457.

(26) Sorensen, J. L.; Auclair, K.; Kennedy, J.; Hutchinson, C. R.; Vederas, J. C. Transformations of cyclic nonaketides by *Aspergillus terreus* mutants blocked for lovastatin biosynthesis at the *lovA* and *lovC* genes. *Org. Biomol. Chem.* **2003a**, *1*, 50-59.

(27) Sorensen, J. L.; Vederas, J. C. Monacolin N, a compound resulting from derailment of type I iterative polyketide synthase function *en route* to lovastatin. *Chem. Commun.* **2003b**, *13*, 1492-1493.

(28) Tobert, J. A. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* **2003**, *2*, 517-526.

(29) Yu, J.; Chang, P.-K.; Ehrlich, K.C.; Cary, J.W.; Bhatnagar, D.; Cleveland, T.E.; Payne, G.A.; Linz, J.E.; Woloshuk, C.P.; Bennett, J.W. Clustered pathway genes in aflatoxin biosynthesis. *Appl. Environ. Microbiol.* **2004**, *70*, 1253-1262.

Table 1. Summary of genes identified in BAC mps01 obtained from *M. pilosus* BCRC38072

mok genes	Amino acids[a]	Putative molecular weight (kDa)	Proposed function[b]	Homologous *lov* gene[c]	Protein similarity (%)[d]	Homologous *mlc* gene[e]	Protein similarity (%)[d]
mokA	3075	338.1	polyketide synthase	*lovB*	76	*mlcA*	66
mokB	2547	278.3	polyketide synthase	*lovF*	73	*mlcB*	61
mokC	524	60.6	P450 monooxygenase	*lovA*	85	*mlcC*	67
mokD	263	28.9	Oxidoreductase	*lovG*	67	*mlcF*	53
mokE	360	38.9	Dehydrogenase	*lovC*	81	*mlcG*	70
mokF	413	46.8	Transesterase	*lovD*	74	*mlcH*	63
mokG	1052	113.0	HMG-CoA reductase	*lvrA*	69	*mlcD*	39
mokH	455	49.4	Transcription factor	*lovE*	54	*mlcR*	49
mokI	543	57.5	Efflux pump	*lovI*	81	*mlcE*	68

[a]The deduced amino acid sequences were determined from cDNA sequences.

[b]The proposed gene functions were based on their homology to proteins in GenBank database.

[c]The lovastatin biosynthetic gene cluster.

[d]Similarity was obtained by alignment using VectorNTI 9.0 (InforMax) software.

[e]The compactin biosynthetic gene cluster.

TABLE 2. Real-time RT-PCR analysis of relative expression of transformant T-mokH1 to *M. pilosus* BCRC38072

Gene cluster[a]	*mokA*	*mokB*	*mokC*	*mokD*	*mokE*	*mokF*	*mokG*	*mokI*
ΔC_T (38072 mk gene-18S rRNA)	2.78 ± 0.82	3.01 ± 0.80	4.14 ± 0.21	3.40 ± 0.72	4.28 ± 0.90	4.57 ± 0.72	4.83 ± 1.28	4.35 ± 1.41
ΔC_T (T-mokH1 mk gene-18S rRNA)	2.11 ± 0.07	2.78 ± 0.56	2.92 ± 0.15	2.68 ± 0.37	2.89 ± 0.32	3.27 ± 0.39	3.26 ± 0.30	3.43 ± 2.02
$\Delta\Delta C_T$ (ΔC_T T-mokH1- ΔC_T 38072)	-0.61 ± 0.07	-0.23 ± 0.56	-1.22 ± 0.15	-0.72 ± 0.37	-1.39 ± 0.32	-1.29 ± 0.39	-1.57 ± 0.30	-0.92 ± 2.02
T-mokH1/BCRC38072 (mean fold change)[b]	1.6	1.2	2.3	1.6	2.6	2.4	3.0	1.9

[a] The values were based on three determinations.

[b] Relative expression was presented as fold difference.

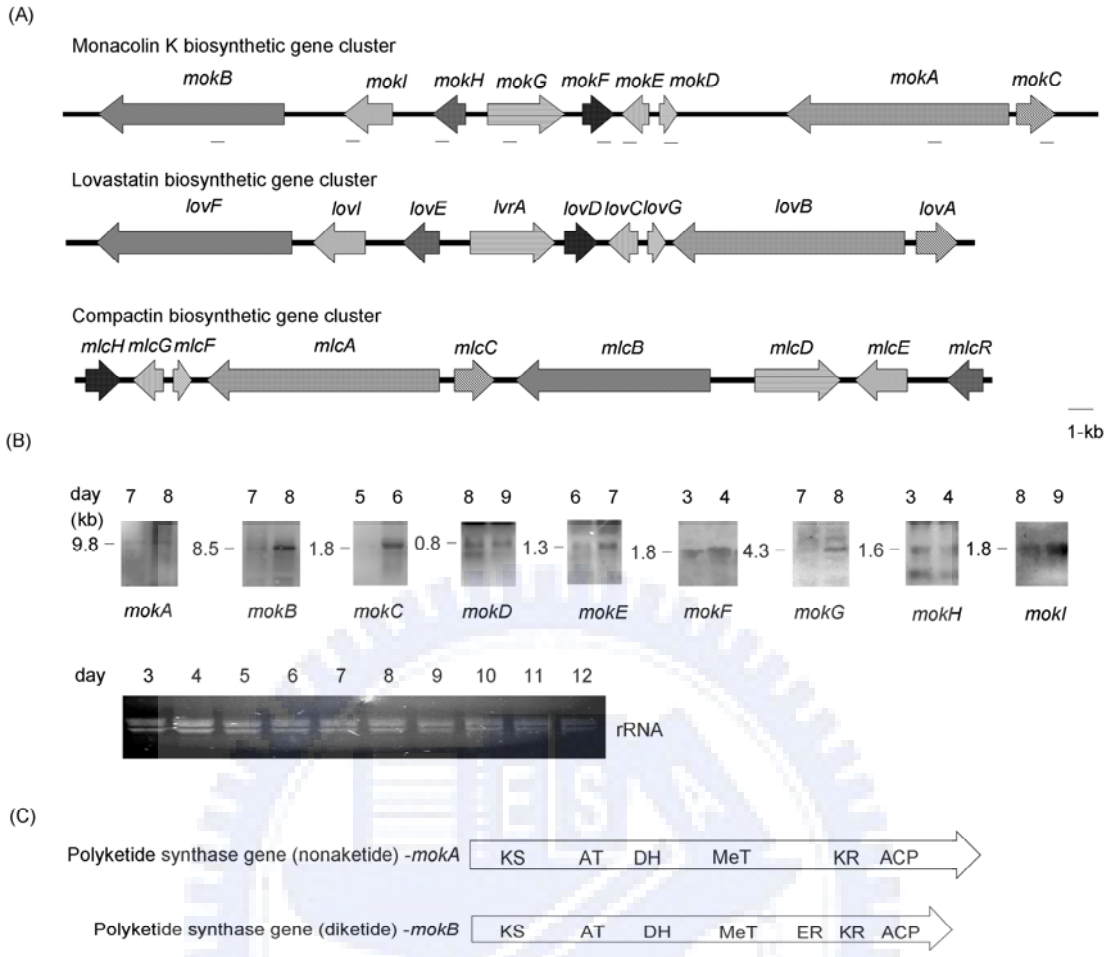


Figure 1. Identification of the monacolin K biosynthetic gene cluster of *Monascus pilosus* BCRC38072. (A) The genes involved in the biosynthesis of monacolin K. For proposed functions of assigned ORFs, see **Table 1**. The small black bars indicated the probes for Northern hybridization analysis. The lovastatin biosynthetic gene cluster in *A. terreus* was obtained from the GenBank database using the following accession numbers: AF141924, AF151722, and AF141925. The compactin biosynthetic gene cluster in *P. citrinum* was obtained from the GenBank database using the following accession number: AB072893. (B) Northern hybridization analyses. Total RNA isolated after 12 days of cultivation was blotted. Total RNA (6 μg per lane) was separated on 0.8 % agarose gels by electrophoresis. The size of each transcript was estimated by comparison with the marker of known

size. (C) Arrangement of functional domains of genes encoding PKSs in the biosynthesis of monacolin K. KS, keto-synthase; AT, acyltransferase; DH, dehydratase; MeT, methyltransferase; ER, enoyl reductase; KR, keto-reductase; ACP, acyl carrier protein.

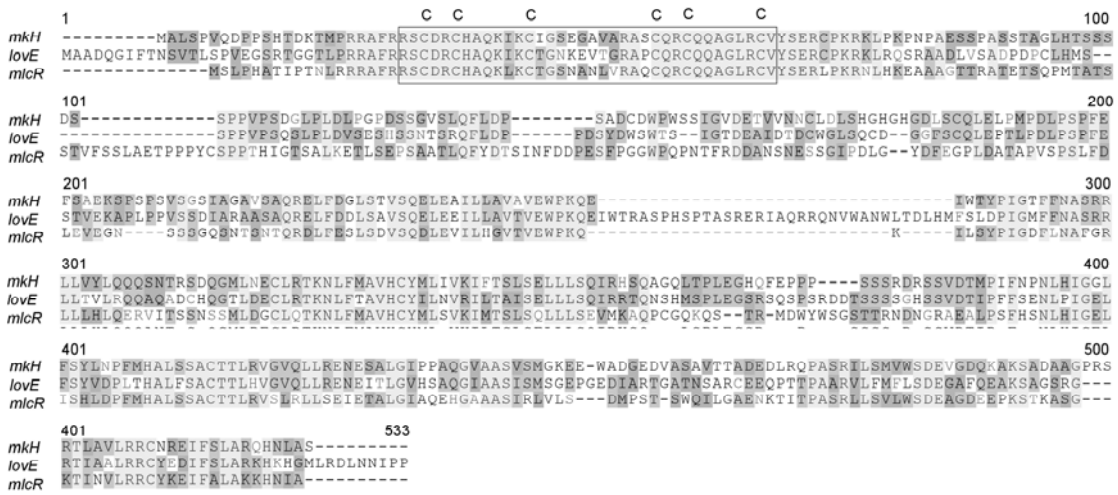


Figure 2. Deduced amino acid sequences alignment of transcription factors from the *mokH*, *lovE*, and *mlcR* genes. The cysteine-rich nucleotide-binding domain represented a Zn_2Cys_6 type zinc finger with the consensus sequence $CX_2CX_6CX_{11}CX_2CX_6C$ shown boxed.

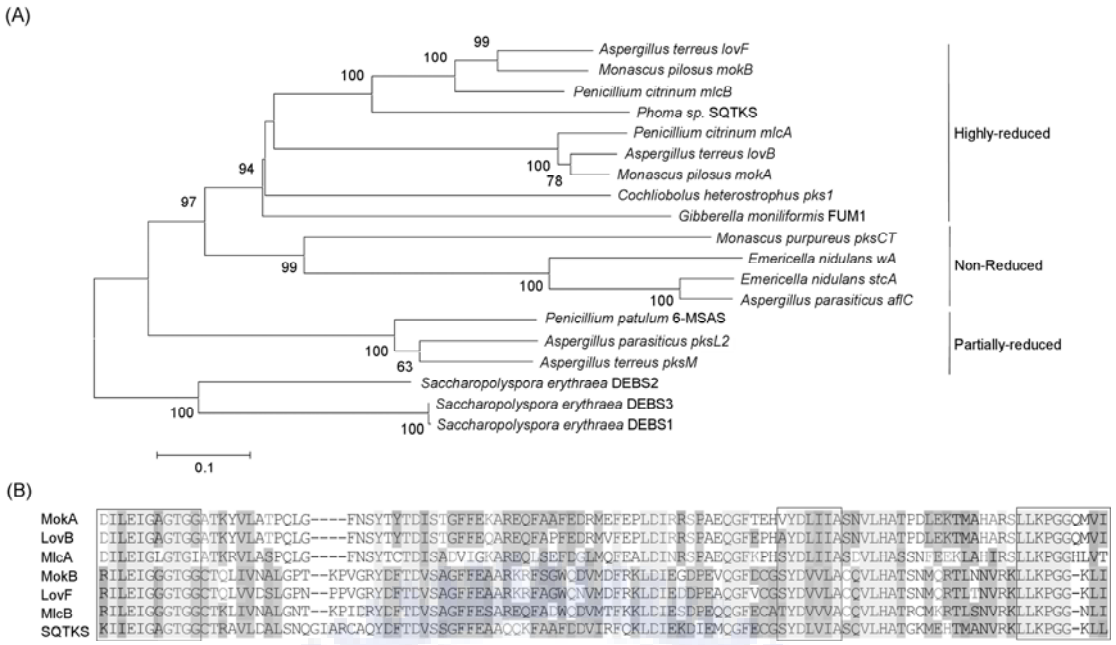
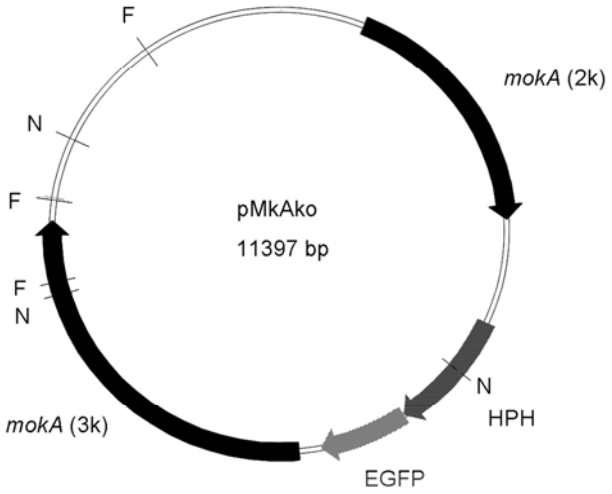


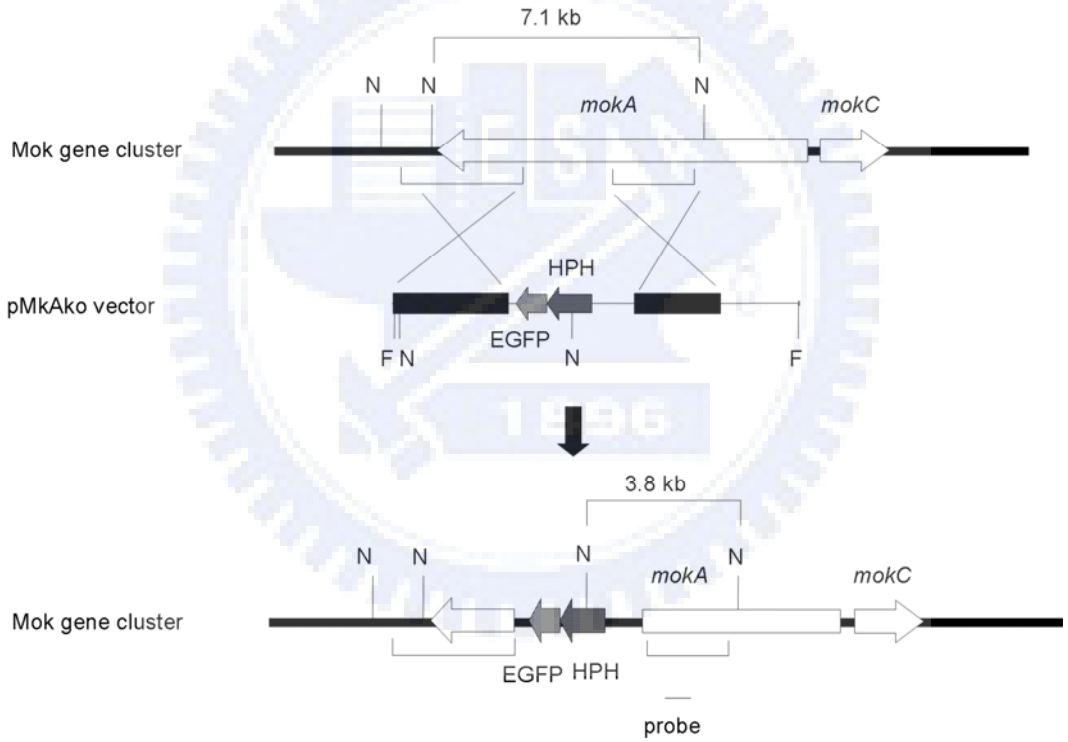
Figure 3. (A) Phylogenetic tree of PKSs from *M. pilosus* BCRC38072 and various organisms. The phylogeny of PKSs based on the conserved keto-synthase domains described by Kroken *et al.* (2003) was constructed and rooted using keto-synthase domains of *Saccharopolyspora erythraea* DEBS (X56107 and X62569). Accession numbers for the polyketide synthase genes were used as follows: *Aspergillus terreus lovF* (AF141925), *Aspergillus terreus lovB* (AF151722), *Penicillium citrinum mlcA*, *mlcB* (AB072893), *Phoma* sp. SQTKS (AY217789), *Cochliobolus heterostrophus pks1* (U68040), *Gibberella moniliformis* FUM1 (AF155773), *Monascus purpureus pksCT* (AB167465), *Emericella nidulans wA* (X65866), *Emericella nidulans stcA* (AAC49191), *Aspergillus parasiticus aflC* (AY371490), *Penicillium patulum* 6-MSAS (X55776), *Aspergillus parasiticus pksL2* (U52151), and *Aspergillus terreus pksM* (U31329). Bootstrap values were shown in the nodes according to 1000 replications. Only bootstrap values >50 % are shown. The tree was constructed by the neighbor-joining method (Saitou and Nei, 1987). (B) Comparison of the methyltransferase (MeT) domain. The three MeT consensus motifs were shown

boxed. The conserved residues of MeT are described by Kagan and Clarke (1994).

(A)


F
mokA (2k)
N
F
pMkAko
11397 bp
F
N
N
HPH
mokA (3k)
EGFP

(B)


7.1 kb
N N N
mokA
mokC
Mok gene cluster
HPH
pMkAko vector
EGFP
F N N F
3.8 kb
N N N N
mokA
mokC
Mok gene cluster
EGFP HPH
probe

(C)

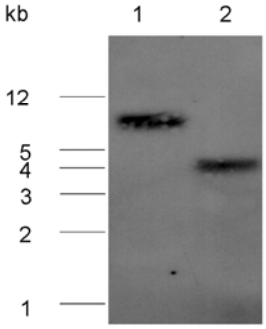

kb
1
2
12
5
4
3
2
1

Figure 4. (A) Plasmid map of pMkAko for targeted gene disruption of *mokA*. (B) Disruption of the *mokA* gene in *M. pilosus* BCRC38072. The strategy for disrupting *mokA* gene was done by homologous recombination. A pMkAko vector containing the fusion protein of the hygromycin B resistance gene (HPH) with enhanced green fluorescent protein (EGFP) was flanked at the 5' site by 2.0 kb (mokA2k) and at the 3' site by 3.1 kb (mokA3k). pMkAko was linearized at the *Fsp*I sites and transformed into *M. pilosus* BCRC38072. The homologous recombination event between the *Monascus* genome and the *Fsp*I-digested pMkAko fragment resulted in a truncated open reading frame for *mokA* gene. (C) Southern hybridization analysis of disruption of *mokA* gene in the genomes of *M. pilosus* BCRC38072 (lane 1), and disruptant BCRC38135 (lane 2) hybridized with the probe indicated by a small black bar. The abbreviation F indicates an *Fsp*I restriction enzyme site, and N indicates an *Nde*I restriction enzyme site.

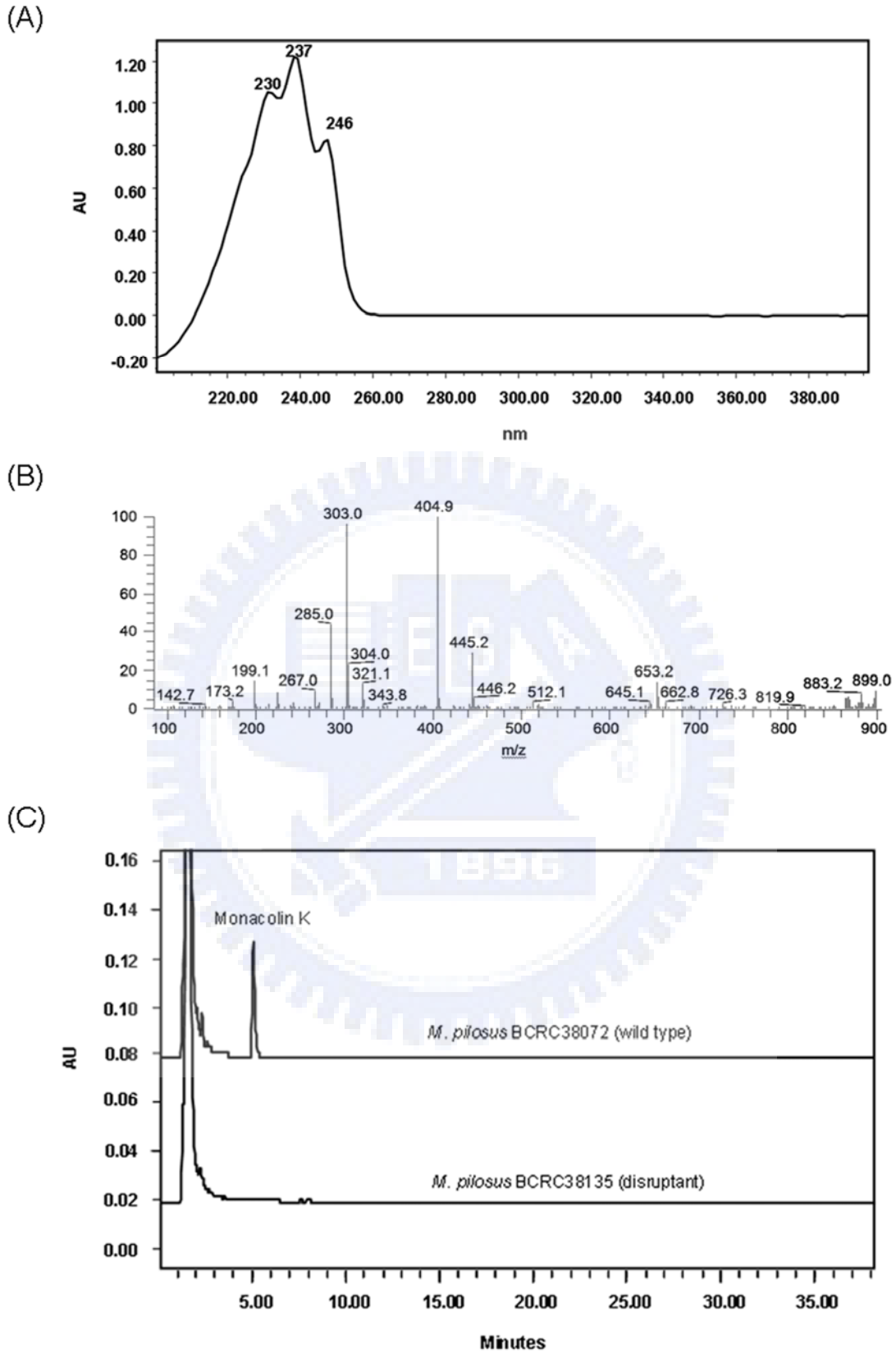


Figure 5. Identification of monacolin K produced by *M. pilosus* BCRC38072. (A) UV

light absorption spectrum of monacolin K between the wavelengths of 210~380 nm. (B) Liquid chromatography-mass spectroscopy analysis of monacolin K. m/z, mass-to-charge ratio. (C) HPLC analysis of monacolin K produced from *M. pilosus* BCRC38072 and BCRC38135 on the eighth day of cultivation.

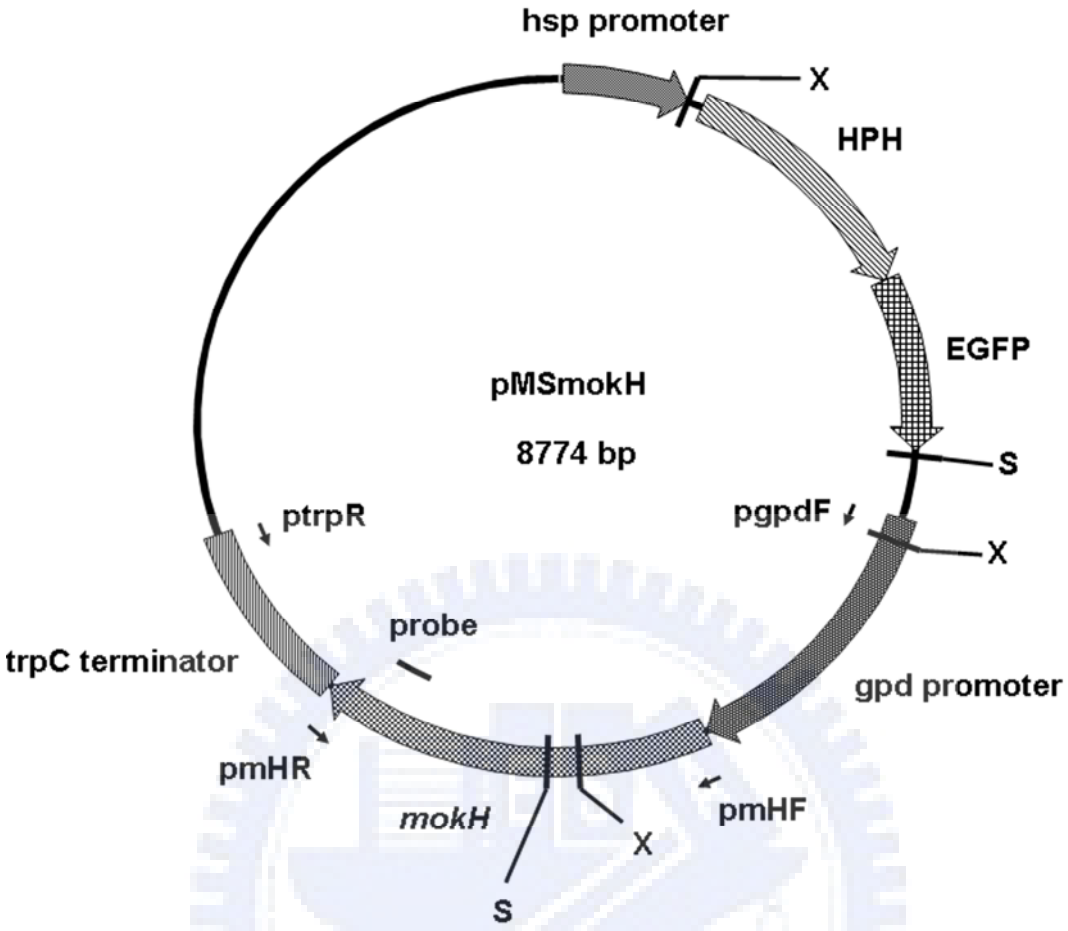


Figure 6. *mokH* gene expression vector pMSmokH for genetic transformation in *M. pilosus* BCRC38072. The abbreviation S indicated the sites of *Sal*I restriction enzyme and X indicated the sites of *Xho*I restriction enzyme.

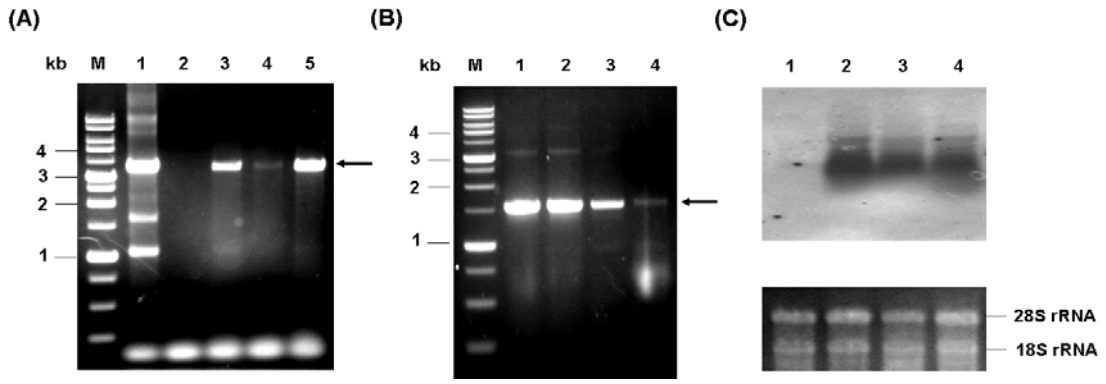


Figure 7. (A) Analysis of PCR of the *mokH* gene containing *gpd* promoter and *trpC* terminator with the primer set of pgpdF and ptrpR in the DNA of pMSmokH (lane1), *M. pilosus* BCRC38072 (lane 2), T-mokH1 (lane 3), T-mokH2 (lane 4) and T-mokH3 (lane 5). (B) Analysis of nested-PCR of the *mokH* gene with the primer set of pmHF and pmHR in the DNA of pMSmokH (lane1), T-mokH1 (lane 2), T-mokH2 (lane 3) and T-mokH3 (lane 4). (C) Analysis of Northern hybridization of the *mokH* gene in the DNA of pMSmokH (lane1), T-mokH1 (lane 2), T-mokH2 (lane 3) and T-mokH3 (lane 4) hybridized with the *mokH* probe.

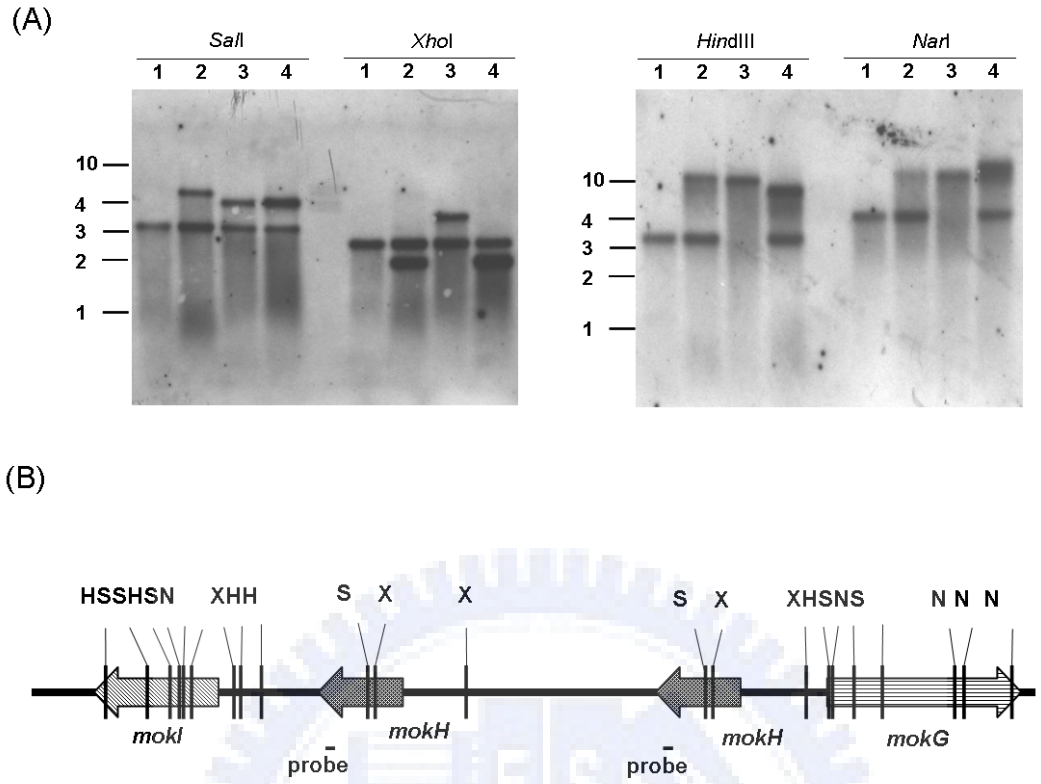


Figure 8. (A) Southern hybridization analyses of copy number of the *mokH* gene in the genomes of *M. pilosus* BCRC38072 (lane 1), T-mokH1 (lane 2), T-mokH2 (lane 3) and T-mokH3 (lane 4) hybridized with the *mkH* probe. (B) The extra copy of *mokH* gene inserted into monacolin K gene cluster in the transformant T-mokH2. S: *Sal*I restriction enzyme; X: *Xho*I restriction enzyme; H: *Hin*dIII restriction enzyme, and N: *Nar*I restriction enzyme.

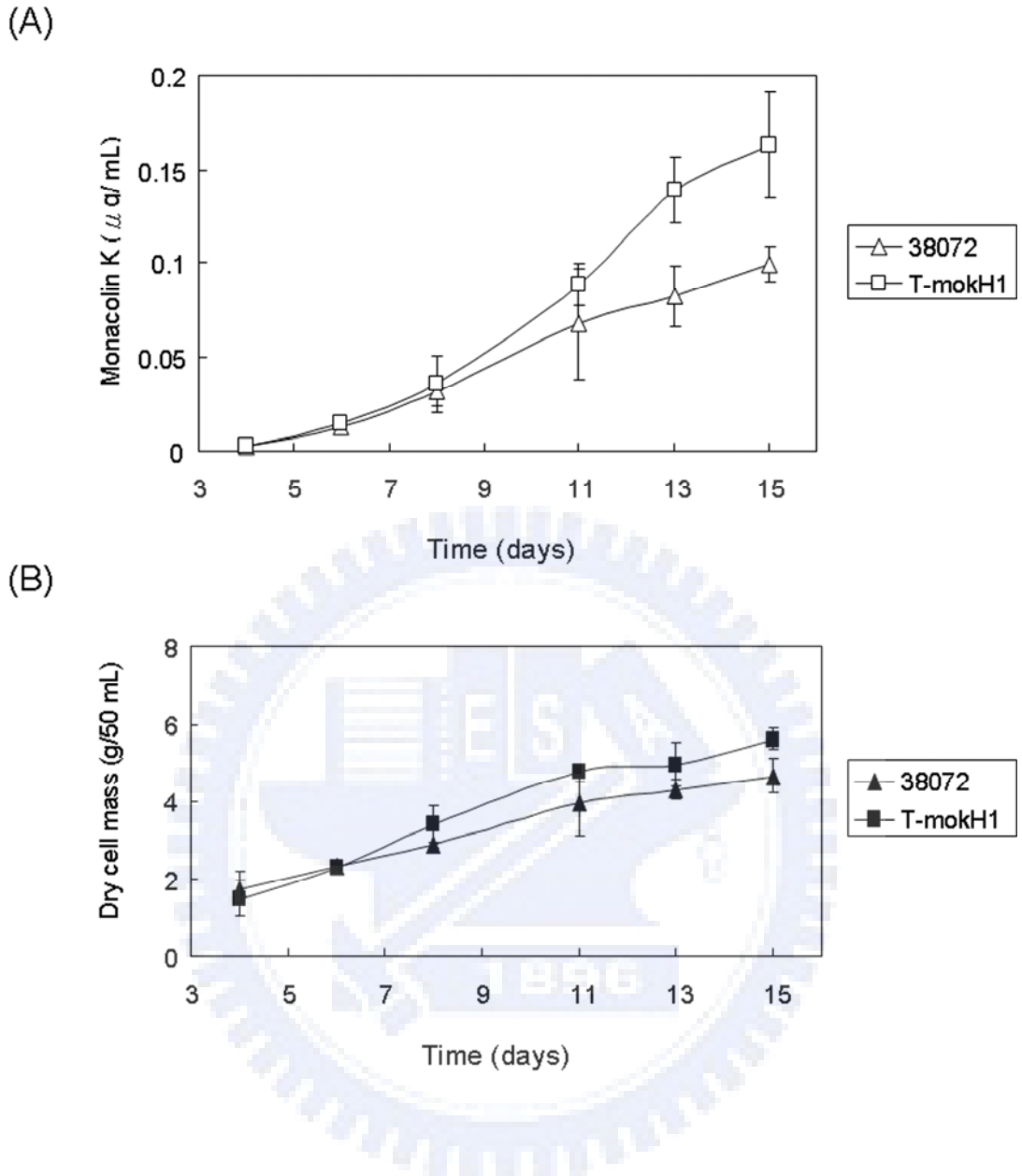


Figure 9. Submerged cultures of the wild type BCRC38072 and the T-mokH1 transformant were incubated on a 200 rpm rotary shaker for 15 days at 25 °C: (△) monacolin K production of the wild type; (□) monacolin K production of the transformant; (▲) dry cell mass of the wild type; (■) dry cell mass of the transformant.

Chapter 3

Overexpression of *mokA* Encoding the Polyketide Synthase of *Monascus pilosus* in *Escherichia coli*

Abstract

The *mokA* gene encoded nonaketide synthase. The *sfp* gene obtained from *Bacillus subtilis* is a phosphopantetheinyl transferase required to convert the expressed apo-PKS to its holo form. *mokA* and *sfp* were coexpressed in *Escherichia coli*. Interestingly, an unexpected product, anthranilic acid which is the precursor of tryptophan, was observed, and the transcript of *trp* operon involved in the biosynthesis of tryptophan from *E. coli* harboring the MkA and Sfp was 250-fold higher than the control. These results revealed that that expression product of polyketide synthase (*mokA*) in *E. coli* was responsible for the elicitor to improve the activity of anthranilate synthase. The high activity of *trpE* gene would lead to anthranilate accumulation.

Introduction

Bacteria produced polyketide are known to be classified into type I and type II systems according to the biosynthesis of polyketide. Modular system (type I) contains many active sites, called a module consisting of motifs, including keto-synthase (KS), acyltransferase (AT), keto-reductase (KR), dehydratase (DH), enoyl rductase (ER) and acyl carrier protein (ACP) etc. such as antibiotics erythromycin (Cane et al., 1998). Iterative system (type II) containing each active site is encoded by separate enzymes. There is a significant characterization of type II polyketide compound with the aromatic group such as antitumor aclacinomycins (Räty et al., 2002). The biosynthetic pathway of polyketides produced by fungi is different from bacteria. Fungal PKSs defined as iterative type I make polyketides similar to bacterial type II, but active sites of PKS are encoded by one enzyme instead of separate enzymes. Furthermore, fungal PKSs are classified into three types according to the reduction of polyketide structures (Nicholson et al., 2001).

Only a few attempts have been made at heterologous expression of fungal polyketide synthase. The 6-methylsalicylic acid synthase (6-MSAS) from *Penicillium patulum*, which involves in 6-methylsalicylic acid biosynthesis, is respectively expressed in *Streptomyces coelicolor*, *E. coli* and *Saccharomyces cerevisiae* (Bedford et al., 1995; Kealey et al., 1998). Lovastatin nonaketide synthase (LNKS) from *Aspergillus terreus*, which involves in lovastatin biosynthesis, is expressed in *Aspergillus nidulans* (Kennedy et al., 1999). PKS1 from *Colletotrichum lagenarium*, which involves in melanin biosynthesis, is expressed in *Aspergillus oryzae* (Fujii et al., 1999). Also, Squalestatin tetraketide synthase (SQTKS) from *Phoma* sp., which involves in squalestatin S1 1 (also known as zaragozic acid A) biosynthesis, is expressed in *Aspergillus oryzae* (Cox et al., 2004).

The utility of *E. coli* as hosts of heterologous polyketide synthase expression has been studied (Kealey et al., 1998; Gokhale et al., 1999). Since a lack of posttranslational modification in *E. coli* prevents heterologous expression of functional polyketide synthase, it is reasonable to coexpress PKS with 4'-phosphopantetheine transferase (PPTase) to produce holo-ACP (acyl carrier protein) domain of PKS for polyketide production (Mootz et al., 2001). In the present study, coexpression of fungal polyketide synthase (*mokA*) from *M. pilosus* BCRC38072 and P-pant transferase (*sfp*) from *B. subtilis* was carried out in *E. coli* to produce novel polyketide.

Materials and Methods

Strain and growth conditions

M. pilosus BCRC38072, which is a moncolin K producing strain isolated from Taiwan, was used in this study. The strain was incubated on YM (DIFCO 271120, Detroit, Michigan) agar for one week, and spore suspensions were obtained by washing cultured YM agar plates with distilled water. Mycelium was harvested by incubating 12 days at 25^{o}C with constant agitation (250 rpm) in liquid medium (7% glycerol, 3% glucose, 3% MSG, 1.2% polypetone, 0.2% $NaNO_3$ and 0.1% $MgSO_4.7H_2O$).

Nucleic acid manipulations

Fungal genomic DNA was isolated according to the method developed by Bingle et al. (1999) and Bacterial Genomic DNA Mni-prep Kit (V-gene, Zhejiang, China) was used to isolate the genomic DNA of *B. subtilis* 168. Colony hybridization, southern hybridization and northern hybridization were performed using DIG system

(Roche Diagnostics, Mannheim, Germany). Total RNA of *M. pilosus* BCRC 38072 was isolated by TRIzol reagent (Invitrogen, Carlsbad, CA). All other DNA manipulations were performed as described in Sambrook et al. (1989).

Construction of *E. coli* expression vector

The partial cDNA fragment of *mokA* gene, 6.5 kb cDNA amplified from the first strand cDNA of *M. pilosus* BCRC38072 with forward primer p3: 5'-**CCATCAAGGATGCTCTGTCG**-3' and reverse primer p4: 5'-**TCAAGCCAACTTCAACGCGG**-3', was ligated with pGEM-T vector (Promega, Madison, WI) to obtain the *Eco*RI—*Not*I fragment by restricted reaction. The fragment was further ligated with pET30 to give pETmkA1. One set of oligonucleotide primers (p1: 5'-GGAATTC**ATGTACGTAGGACGCATTGGTGC** -3' containing 23 bases complementary to the 5' *mokA* gene and introducing *Eco*RI restriction site; and p2: 5'- **TCGCGAGGACGGACAAAGTT**-3') was designed to amplify the partial 3.0 kb cDNA which was restricted to obtain the *Eco*RI—*Eco*RI fragment. The 3.0 kb *Eco*RI—*Eco*RI cDNA fragment was ligated with pETmkA1 to give pETmkA. The *sfp* gene isolated from from genomic DNA of *B. subtilis* 168 was excised as a 690 bp *Bam*HI—*Not*I fragment from amplification with primers 5'-CGGGATCCC**ATGAAGATTTACGGAATTTA** -3' and 5'-ATAGTTTAGCGGCCGC**TTATAAAAGCTCTTCGTACG** -3'. The *Bam*HI—*Not*I *sfp* gene was ligated with pCDF-1b to give pCDFsfp.

Co-expression of *mokA* and *sfp* in *E. coli*

E. coli^{star} (DE3) (Invitrogen, Carlsbad, CA) was cotransformed with both pETmkA and pCDFsfp and inoculated with a single colony for culture of overnight. The transformant was transferred into ATCC medium 765 and grown until OD_{600}

reached 0.4-0.5 at 250rpm and 37°C. These cultures were cooled for 10min at 4°C and induced by addition of isopropyl β-D-thiogalactoside (IPTG) to 0.1 mM (final concentration). Sodium propionate was added to 2.25mM (final concentration) and the cultures were incubated for 48 hr at 20°C. Protein expression was monitored by SDS-PAGE (7.5% gels) of total cellular protein and soluble protein followed by Colloidal Coomassie staining.

Identification and characterization of compound

Aliquots of *E. coli* culture were cleared of cells by centrifugation, and supernatants were extracted with ethyl acetate. The extract was dried, resuspended in methanol, and analyzed by HPLC. HPLC was performed on a Waters system fitted with an Inertsil ODS-3 column (1.8 x 250mm). The HPLC parameters were as follows: solvent A, 0.1% acetic acid in water; solvent B, acetonitrile; gradient, 5% B to 10% B in 20 min, 10% B to 30% B in 40 min, 30% B to 40% B in 10 min, 40% B to 50% B in 20 min, then to 100% in 30 min; flow rate, 1 ml/min; detected by UV spectroscopy. The product was analyzed by TLC using silica plates and 3:1 hexan:ethyl acetate mixture as mobile phase. The structure of anthranilic acid was confirmed by EI-MS spectrometry, ^{1}H and ^{13}C NMR spectroscopy.

Quantitation of trptophan related gene expression by real-time RT-PCR

E. coli was harvested after IPTG induction for 48 hr at 20°C. Total RNA was prepared by TRIzol reagent (Invitrogen, Carlsbad, CA) and treated with 2 U of RNase-free DNase for 1 h. First strand cDNA was synthesized by ImProm-II™ Reverse Transcription System with random hexamers (Promega, Madison, WI) at 42°C for 1 h and used as the template for real-time PCR. Real-time PCR was carried out using the Absolute SYBR Green ROX Mix (Abgene, Epsom, UK) in a final

volume of 25 μl for each reaction in the ABI Prism 7000 sequence detection system (Applied Biosystems, Foster City, CA). The thermal cycling parameters consisted of an initial heating at 95°C for 15 min, denaturing at 95°C for 10 s, and annealing and extension at 60°C for 1 min, with an amplification of 40 cycles. The primer set of *E. coli* 16S rRNA designed by the PrimerExpress software was used as a positive control for normalization. RNA samples without a reverse transcriptase step were also analyzed to determine genomic DNA contamination. The primer sets of *trpD* and *trpE* genes were trpDf: CAGGCGATTGTCGAAGCTTAC, trpDr: TGCCAACCAGCGAGTGATAAC; trpEf: GGCGGGATTTGAAGATTTACC, and trpEr: GGCCTGAATACGGGTGCTTT. Measurements of gene expression were performed three times and the mean of these values was used for analysis. The relative gene expression level was calculated by the $2^{-\Delta\Delta CT}$ method as described in Applied Biosystems User Bulletin No. 2.

Results

Expression of Polyketide Synthase in *E. coli*

Based upon earlier polyketide biosynthetic schemes, lovastatin nonaketide synthase (LNKS) has successfully expressed in *A. nidulans*, and produces two novel polyketides, hexaketide pyrone (4-Hydroxy-6-[(1*E*,3*E*,5*E*)-1-methylhepta-1,3,5-trien-1-yl]-2-pyrone) and heptaketide pyrone (4-Hydroxy-6-[(1*E*,3*E*,5*E*,7*E*)-3-methylnona-1,3,5,7-tetraen-1-yl]-2-pyrone) (**Figure 1**) (Kennedy et al., 1999). Because *E. coli* does not produce polyketides, we initiated our attempts toward heterologous polyketide synthesis by expression of MokA involved in the biosynthsis of nonaketide skeleton in *E. coli*^{star} (DE3). Initially, reverse-transcription PCR was carried out to

obtain two cDNA fragments of MokA. Two cDNA fragments of *mokA* gene were further reassembled and reconstructed as an expression plasmid pETmkA. Fungal PKS gene expression was analyzed at lower temperature (20^{o}C) and lower dose IPTG (100 μM) for post-induction. Growth of *E. coli* harboring the MokA and Sfp expression plasmids, pETmkA and pCDFsfp, for 48 hr resulted in expression of two proteins that migrated on SDS/PAGE gels at 342 and 29 kDa, the predicted sizes of MokA and Sfp (**Figure 2**). The 342 kDa protein was not expressed in *E. coli* containing control plasmid pET30. The coexpression vectors used in this study were engineered with an NH_2-terminal His-tag. Recombinant fusion proteins were purified by Ni^{+}-resin column.

The production of compound by *E. coli* harboring both MokA and Sfp expression plasmids was monitored by HPLC of cell-free supernatants. The transformant showed the present of a new compound in the HPLC trace. This compound was isolated and purified by repeated TLC. The compound showed UV light induced blue fluorescence by TLC analysis. Detection was performed by measuring the UV absorption at wavelengths of 218 nm and 330 nm, and the retention time was 51 min by HPLC analysis (**Figure 3**). Mass spectral analysis of the product gave a 137 atomic mass units. ^{1}H and ^{13}C NMR analyses confirmed the product: (^{1}H NMR [CD_3Cl, 400 Mhz] δ [ppm] 6.66~6.69 [2H, m, H-4, 3], 7.31 [1H, td, J = 8.0, 1.6 Hz, H-2], 7.92 [1H, dd, J = 8.0, 1.6 Hz, H-1] , with tetramethylsilane as the internal reference). (^{13}C NMR [CD_3Cl, 500 Mhz] δ [ppm] 110.1, 117.1, 117.4, 132.8, 135.7, 151.7 and 173.7, with tetramethylsilane as the internal reference). Full structural characterization revealed this compound to be the anthranilic acid by EI-MS, ^{1}H and ^{13}C NMR analyses (**Figure 4**).

Discussion

The development of *E. coli*-derived polyketide marks an important step forward for heterologously produced complex polyketide product from fungi (Kealey et al., 1998). Although it is not presumably efficient catalyzed by the endogenous *E. coli* holo-acyl carrier protein synthase, it is reasonable to coexpress with P-pant transferase to produce a holo-PKS capable of polyketide production (Pfeifer and Khosla, 2001). 6-MSAS from *P. patulum* is the only PKS from the fungus that has been successfully expressed in *E. coli* (Kealey et al., 1998). In this study, the reconstructed 9,381 bp PKS contained full-length cDNA of *mokA* gene, and his-tag based on pET30 vector. The expression plasmid was exploited in the *E. coli* expression system. Interestingly, an unexpected product, anthranilic acid, was found and produced as much as 10mg per liter of culture. Anthranilic acid is catalyzed from chorismate and glutamine by *trpE* gene encoding anthranilate synthase component I which is the first enzyme in the tryptophan biosynthetic pathway. The *trpD* gene encoding anthranilate phosphoribosyltransferase, also called anthranilate synthase component II, converts anthranilte and 5-phosphoribosylpyrophosphate to phosphoribosylanthranilate. Anthranilate synthase activity is subject to feedback inhibition in the present of L-tryptophan (Ito and Yanofsky, 1969). For example, a corn (*Zea mays L.*) mutant that accumulates anthranilic acid is also described. Anthranilate synthetase from mutant plant is 3-40 fold more active than that of normal plant and is resistant to feedback inhibition by tryptophan (Singh and Widholm, 1975). Hence, we pursued quantitatively using real-time PCR for transcripts of *trp* operon. It was shown that the transcript of anthranilate synthetase from *E. coli* harboring the MokA and Sfp expression plasmids pETmkA and pCDFsfp was roughly 250-fold higher than that harboring control plasmids pET30 and pCDF-1b (**Figure 5**). Further study on the

feedback inhibition by tryptohan was proven. When L-tryptophan was added at final concentration of 500 μM after induction of IPTG, there was no phenomenon of anthranilate accumulation. The result demonstrated that high activity of *trpE* gene would lead to anthranilate accumulation.

The shikimic acid pathway in bacteria, fungi, and plants generates the aromatic amino acids phenylalanine, tyrosine, and tryptophan, which leads to the biosynthesis of numerous secondary metabolites (Coggins et al., 2003). The polyketide synthesis of the aromatic polyene macrolide antibiotic candicidin, produced by *Streptomyces griseus*, is found that is inhibited *in vivo* by L-tryptophan, L-phenylalanine, and, to a lesser degree, L-tyrosine. The regulatory mechanism of candicidin biosynthesis is controlled by the primary metabolites, such as aromatic amino acids, involved in the shikimic acid pathway (Gil and Campelo-Diez, 2003). Furthermore, glycosylated p-hydroxybenzoic acid methyl ester (*p*-HBAD) is an important virulene factor of *Mycobacterium tuberculosis*. The product is though to be derived from *p*-hydroxybenzoic acid, the origin of this biosynthetic precursor. Recently, it is found that the gene *Rv2949c*, located in the vicinity of the polyketide synthase gene *pks15/1*, is involved in the conversion of chorismate to *p*-hydroxybenzoate (Stadthagen et al., 2005). These studies indicate that polyketide synthesis is regulated by the mechanisms controlling the biosynthesis of primary metabolites involved in the shikimic acid pathway. However, in this study, expression product of polyketide synthase (*mokA*) in *E. coli* played the role of "elicitor" that up-regulated the anthranilate biosynthesis involved in the tryptophan biosynthetic pathway. Traditionally, employment of pathogenic and non-pathogenic fungal preparations and chemicals, termed as elicitors, thus become one of the most important strategies to improve secondary metabolite production in cell cultures (Radman et al., 2003). Various fungal elicitors are proven to improve indole alkaloid production (ajmalicine,

serpentine and catharanthine) in *Catharanthus roseus* cell suspension cultures (Wang et al., 2006). Cell cultures of *Ruta graveolens* treated with *Rhodotorula rubra* extract (elicitor) reveal that *ASα1* (subunit of anthranilate synthase component I) steady-state mRNA levels are increased 100-fold at 6 hr (Bohlmann et al., 1995). Although there were no expected products identified in cultures of *E. coli*, the interesting finding indicated that expression product of polyketide synthase (*mokA*) regulated anthranilate flux through the anthranilate synthase gene specific to the tryptophan biosynthetic pathway.

The secondary metabolites of polyketides produced by fungi express structural variety and unique characteristics which do not exist in other bacteria (Pfeifer and Khosla, 2001). These characteristics are also expressed in enzyme variety of polyketide synthesis. It is found that the various polyketides, for example, monacolin K, J, L, M, X, citrinin and pigment from *Monascus* are produced by condensation of acetyl CoA catalyzed by polyketide synthase (PKS). This work, which was the first finding that expression product of polyketide synthase in *E. coli* played the role of "elicitor" to improve anthranilic acid production. Only a few heterologous expression experiments with fungal polyketide synthase genes have been reported in *E. coli* (Kealey et al., 1998). In this study, we demonstrated that fungal polyketide synthase combined with genetic engineered heterologously expression was a great potential for discovery of a diverse range of biological activity.

References

(1) Bedford, D.J.; Schweizer, E.; Hopwood, D.A.; Khosla, C. Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2). *J. Bacteriol*. **1995**, *177*, 4544-4548.

(2) Bingle, L. E. H.; Simpson, T. J.; Lazarus, C. M. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet. Biol.* **1999**, *26*, 209-223.

(3) Bohlmann, J.; DeLuca, V.; Eilert, U.; Martin, W. Purification and cDNA cloning of anthranilate synthase from *Ruta graveolens*: modes of expression and properties of native and recombinant enzymes. *Plant J*. **1995**, *7*, 491-501.

(4) Cane, D.E.; Walsh, C.T.; Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science*. **1998**, *282*, 63-68.

(5) Coggins, J.R.; Abell, C.; Evans, L.B.; Frederickson, M.; Robinson, D.A.; Roszak, A.W.; Lapthorn, A.P. Experiences with the shikimate-pathway enzymes as targets for rational drug design. *Biochem Soc Trans*. **2003**, *31*, 548-552.

(6) Cox, R.J.; Glod, F.; Hurley, D.; Lazarus, C.M.; Nicholson, T.P.; Rudd, B.A.M.; Simpsom, T.J.; Wilkinson, B.; Zhang, Y. Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis. *Chem. Comm*. **2004**, *20*, 2260-2261.

(7) Fujii, I.; Mori, Y.; Watanabe, A.; Kubo, Y.; Tsuji, G.; Ebizuka, Y. Heterologous expression and product identification of *Colletotrichum lagenarium* polyketide synthase encoded by the PKS1 gene involved in melanin biosynthesis. *Biosci Biotechnol Biochem*. **1999**, *63*, 1445-1452.

(8) Gil, J.A.; Campelo-Diez, A.B. Candicidin biosynthesis in *Streptomyces griseus*. *Appl Microbiol Biotechnol*. **2003**, *60*, 633-642.

(9) Gokhale, R.S.; Tsuji, S.Y.; Cane, D.E.; Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science*. **1999**, *284*, 482-485.

(10) Ito, J.; Yanofsky, C. Anthranilate synthetase, an enzyme specified by the tryptophan operon of *Escherichia coli*: Comparative studies on the complex and the subunits. *J Bacteriol*. **1969**, *97*, 734-742.

(11) Kealey, J.T.; Liu, L.; Santi, D.V.; Betlach, M.C.; Barr, P.J. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci. USA.* **1998**, *95*, 505-509.

(12) Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science.* **1999**, *284*, 1368-72.

(13) Mootz, H.D.; Finking, R.; Marahiel, M.A. 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J. Biol. Chem*. **2001**, *40*, 37289-37298.

(14) Nicholson, T.P.; Rudd, B.A.; Dawson, M.; Lazarus, C.M.; Simpson, T.J.; Cox, R.J. Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem. Biol.* **2001**, *8*, 157-178.

(15) Pfeifer, B. A.; Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 106-118.

(16) Radman, R.; Saez, T.; Bucke, C.; Keshavarz, T. Elicitation of plants and microbial cell systems. *Biotechnol Appl Biochem*. **2003**, *37*, 91-102.

(17) Räty, K., Kantola, J., Hautala, A., Hakala, J., Ylihonko, K. & Mäntsälä, P. *Gene*. **2002**, *293*, 115-122.

(18) Sambrook, J.; Fritsch, E.; Maniatis, T. Molecular cloning: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, New York. **1989**.

(19) Singh, M.; Widholm, J.M. Study of a corn (*Zea mays* L.) mutant (blue fluorescent-1) which accumulates anthranilic acid and its beta-glucoside. *Biochem Genet*. **1975**, *13*, 357-367.

(20) Stadthagen, G.; Kordulakova, J.; Griffin, R.; Constant, P.; Bottova, I.; Barilone, N.; Gicquel, B.; Daffe, M.; Jackson, M. p-Hydroxybenzoic acid synthesis in *Mycobacterium tuberculosis*. *J Biol Chem.* **2005**, *280*, 40699-40706.

(21) Wang, W.; Yu, L.; Zhou, P. Effects of different fungal elicitors on growth, total carotenoids and astaxanthin formation by *Xanthophyllomyces dendrorhous*. *Bioresour Technol*. **2006**, *97*, 26-31.

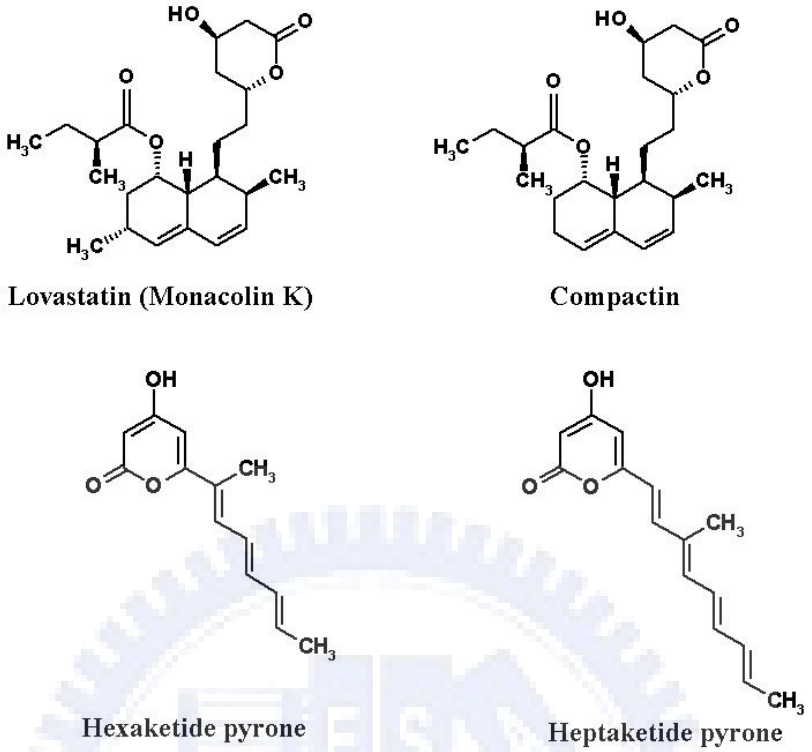


Fig. 1. Structures of natural (lovastatin and compactin) and "unnatural" (hexaketide pyrone and heptaketide pyrone) polyketide products.

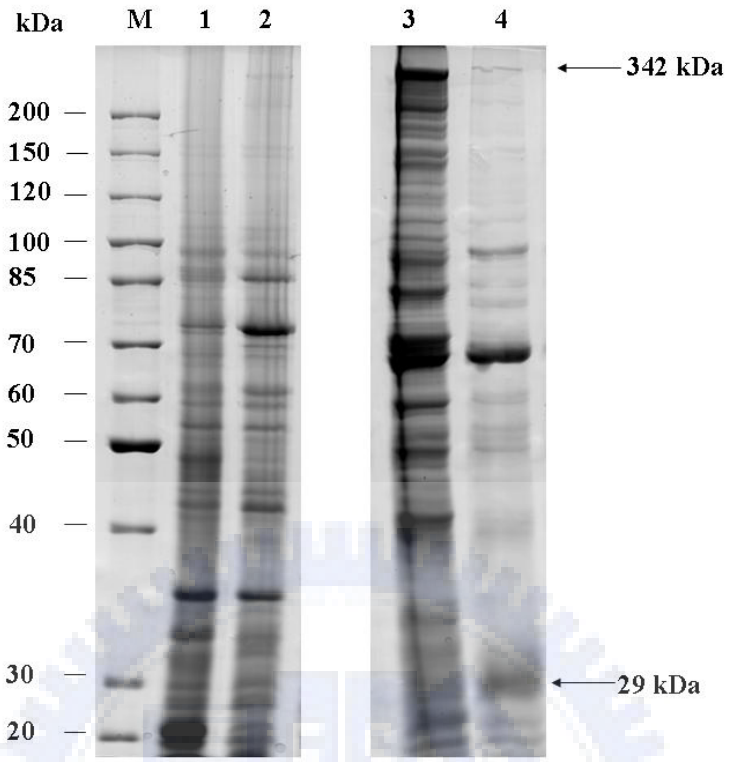


Fig. 2. Expression of MokA and Sfp in *E. coli*. *E. coli* harboring total proteins of control plasmids pET30 and pCDF-1b (lane 1), and that of MokA expression plasmid pETmkA and *sfp* expression plasmid pCDFsfp (lane 2) were grown for 48 hr and boiled for SDS-PAGE. Cells in buffer were lysated and centrifuged to obtain soluble proteins (lane 3). Soluble proteins of expression plasmids pETmkA and pCDFsfp were purified by Ni-column (lane 4).

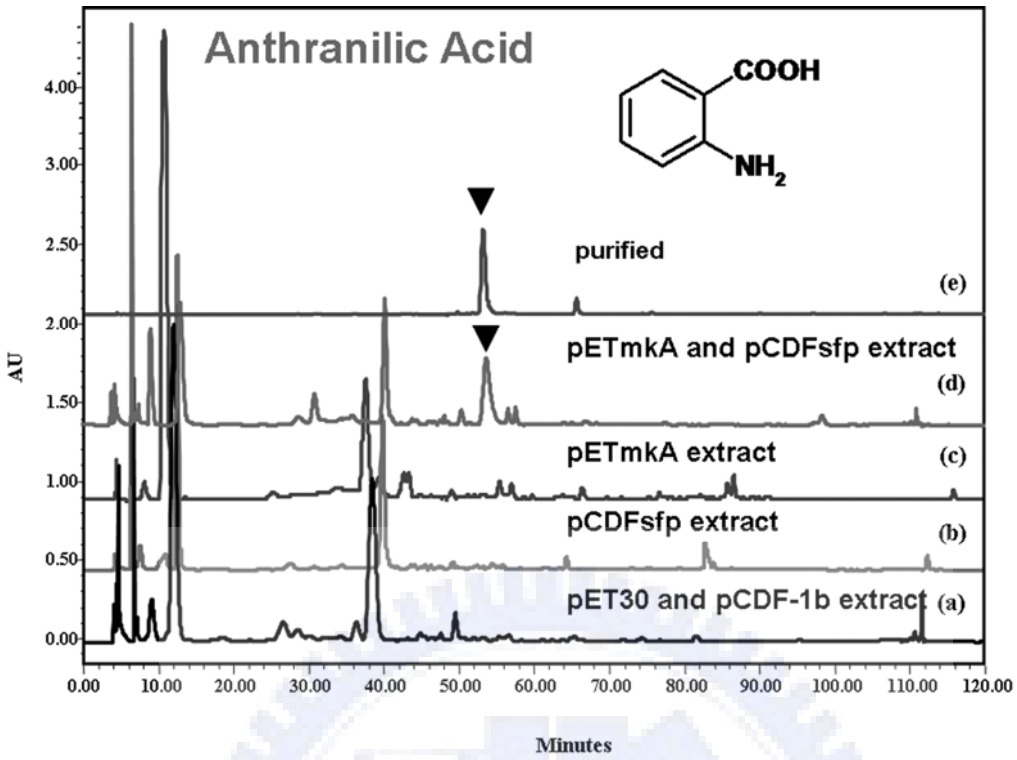
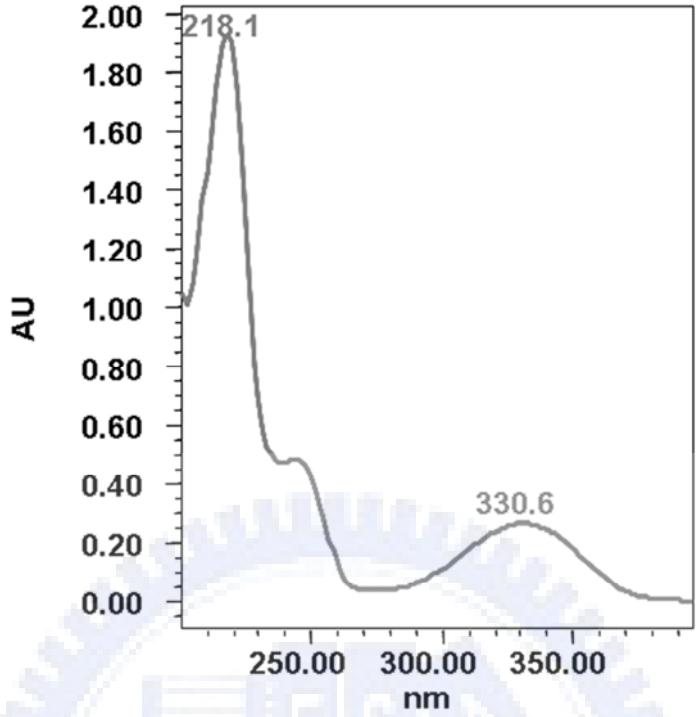
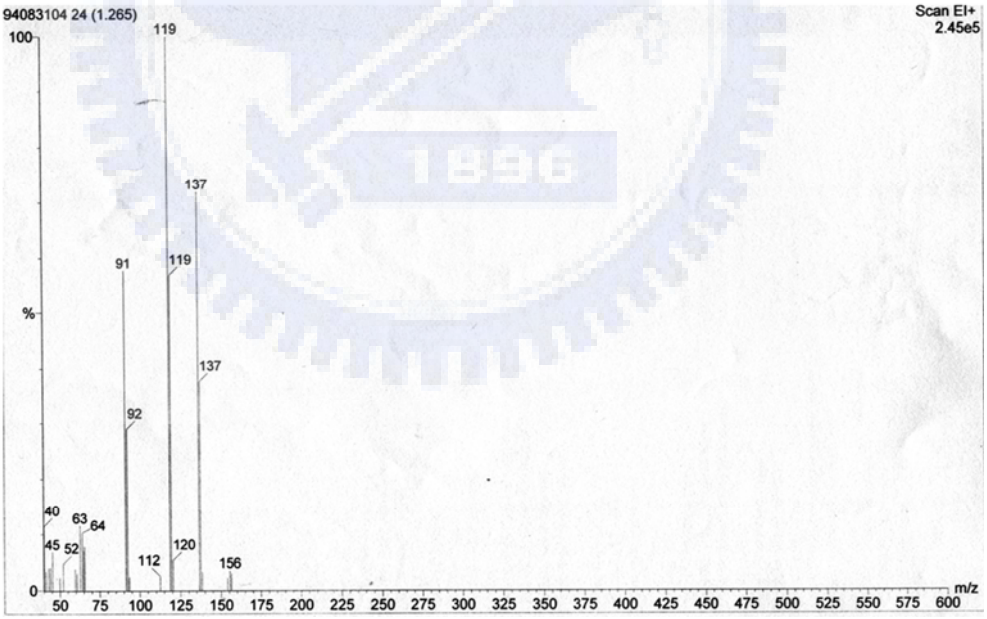


Fig 3. HPLC traces monitored at a UV wavelength of 254 nm including extract from : (a) *E. coli* harboring control plasmids pET30 and pCDF-1b; (b) *E. coli* harboring expression plasmid pETmkA; (c) *E. coli* harboring expression plasmid pCDFsfp; (d) *E. coli* harboring expression plasmids pETmkA and pCDFsfp and (e) purified.

(A)


218.1
330.6
AU
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
250.00
300.00
350.00
nm

(B)


94083104 24 (1.265)
Scan EI+
2.45e5
100
%
0
119
137
91
119
137
92
40
63
64
45
52
112
120
156
50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600
m/z

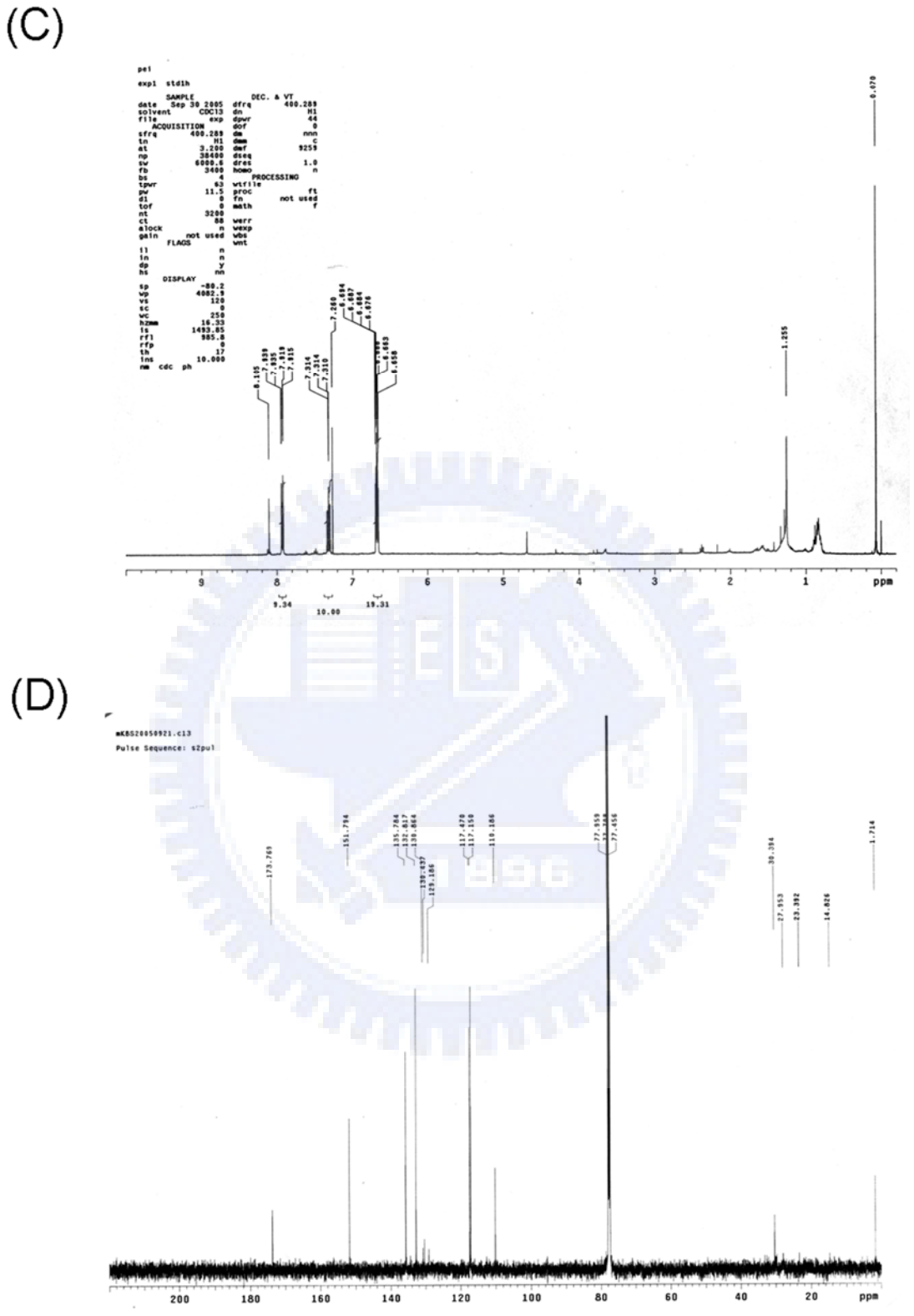


Figure 4. Identification of anthranilic acid produced by *E. coli* with coexpression of pETmkA and pCDFsfp. (A) UV light absorption spectrum of anthranilic acid between the wavelengths of 210~380 nm. (B) Electron ionization-mass spectroscopy analysis

of anthranilic acid. m/z, mass-to-charge ratio. (C) (D) ^{1}H and ^{13}C NMR analyses of anthranilic acid.

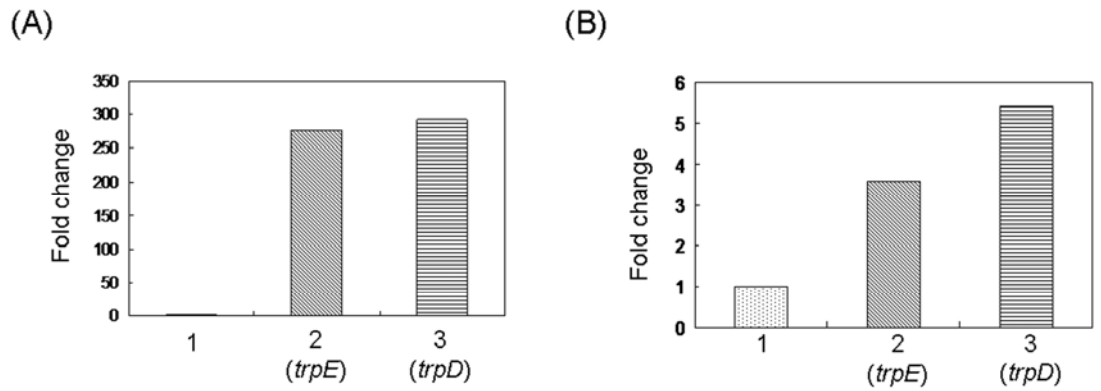


Figure 5. Real-time RT-PCR analysis of relative expression of *E. coli* harboring the MokA and Sfp expression plasmids to control. (A) *trpE* and *trpD* from *E. coli* harboring the MokA and Sfp expression plasmids were 250-fold higher than that harboring control plasmids pET30 and pCDF-1b. (B) *trpE* and *trpD* from *E. coli* harboring the MokA and Sfp expression plasmids were inhibited by adding 500 μM L-tryptophan. 1. Relative expression as control; 2. *trpE* gene expression; 3. *trpD* gene expression.

Chapter 4

Exploring the Distribution of Citrinin Biosynthesis Related Genes among *Monascus* Species

Abstract

Citrinin, a hepato-nephrotoxic compound to humans, can be produced by the food fermentation microorganisms, *Monascus* spp. In this study, we investigated the distribution of mycotoxin citrinin biosynthesis genes in 18 *Monascus* strains. The results show that the acyl-transferase and keto-synthase domains of the *pksCT* gene encoding citrinin polyketide synthase were found in *M. purpureus*, *M. kaoliang* and *M. sanguineus*. Furthermore, the *ctnA* gene, a major activator for citrinin biosynthesis, was found in *M. purpureus* and *M. kaoliang*, but was absent in *M. sanguineus*. The *orf3* gene encoding oxygenase, located between *pksCT* and *ctnA*, was also present in *M. purpureus* and *M. kaoliang*. The *pksCT* gene was highly conserved in *M. purpureus*, *M. kaoliang*, and *M. sanguineus*, while the *ctnA* and *orf3* genes were shown to have highly homologous in *M. purpureus* and *M. kaoliang*. In contrast, the PCR and Southern blot analyses suggest that *pksCT*, *ctnA* and *orf3* were absent or significantly different in *M. pilosus*, *M. ruber*, *M. barkeri*, *M. floridanus*, *M. lunisporas*, and *M. pallens*. A citrinin producing phenotype was detected only in *M. purpureus* and *M. kaoliang* using high performance liquid chromatography (HPLC). These results clearly indicate that the highly conserved citrinin gene cluster in *M. purpureus* and *M. kaoliang* carry out citrinin biosynthesis. In addition, according to the phylogenetic subgroups established with the β-tubulin gene, the citrinin gene cluster can group the species of *Monascus*.

Keyword: citrinin, *Monascus*, polyketide, β-tubulin

Introduction

Typically used in fermented foods, the filamentous fungi *Monascus* spp. are known to be the producers of various secondary metabolites with polyketide structures such as pigments, monacolin K, and citrinin (Endo, 1979; Endo et al., 1986; Shimizu et al., 2005). The red pigments produced by the various species of *Monascus* have been used extensively as natural food colorants (Babitha et al., 2007). Known as lovastatin and normally used for reducing serum cholesterol levels in humans, monacolin K was first isolated from the *Monascus ruber* medium (Endo, 1979), although it can also be found in *Aspergillus terreus* (Hendrickson et al., 1999; Kennedy et al., 1999; Tobert, 2003). Citrinin, the hepato-nephrotoxic agent and antibiotic against gram-positive bacteria, has been identified in a variety of fungi such as *Aspergillus* and *Penicillium* spp. (Sweeney and Dobson, 1998; Malmstrøm et al., 2000); the same substance is also found in *Monascus* as monascidin A (Blanc et al., 1995a; Blanc et al., 1995b). It belongs to the group of polyketides synthesized by the iterative type I polyketide synthase (PKS). The biosynthesis of citrinin originates from a tetraketide arising from the condensation of one acetyl-CoA molecule with three malonyl-CoA molecules in *Monascus* (Hajjaj et al., 1999a). Although *Monascus* has been widely used in food fermentation, and has also shown a highly promising application in medicine development, the nephrotoxic and hepatotoxic properties of citrinin limit its application.

The citrinin biosynthetic gene cluster was proposed in *M. purpureus* BCRC33325 (IFO30873) in Shimizu's studies (2005). The polyketide synthases (*pksCT*) and transcriptional activator (*ctnA*) have been proven to be involved in citrinin biosynthesis (Shimizu et al., 2005; Shimizu et al., 2007). The *pksCT* encoding the PKS is responsible for such biosynthesis. As such, the disruption of

pksCT results in the phenotype of lost citrinin productivity in the disruptant (Shimizu et al., 2005). The *ctnA* that encodes the Zn(II)2Cys6 binuclear DNA binding protein is a major activator of citrinin biosynthesis. The *ctnA*-disrupted strain of *M. purpureus* also brings a significant decrease in citrinin production down to a barely detectable level (Shimizu et al., 2007). To characterize the citrinin-produced *Monascus* strains, we analyzed the genotype and phenotype of the citrinin synthesized by the various species of *Monascus*. According to the citrinin biosynthetic gene cluster in *M. purpureus* BCRC33325 (IFO30873) (Shimizu et al., 2005; Shimizu et al., 2007), several primers were designed for polymerase chain reactions (PCR) and Southern hybridizations. Citrinin productivity was further determined through HPLC analysis. Our results indicated that the distribution of citrinin biosynthetic genes was highly related to the phylogenetic group of *Monascus* spp.

Materials and methods

Strains, media and growth conditions

Eighteen strains of *Monascus* listed in **Table 1** were used in this study. All strains were maintained on PDA (DIFCO, Detroit, Michigan) agar for one week, and spore suspensions were obtained by washing the cultured PDA agar plates with distilled water. Mycelia were harvested after incubating for 14 days at 25^{o}C with constant agitation in liquid medium (7% glycerol, 3% glucose, 3% monosodium glutamate, 1.2% polypetone, 0.2% $NaNO_3$, and 0.1% $MgSO_4.7H_2O$).

Genomic DNA isolation

Monascus genomic DNA was extracted according to the method developed by

Bingle et al. (1999). Approximately 0.5 g (squeezed wet weight) of frozen mycelia was ground to a fine power under liquid nitrogen using mortar and pestle. Proteins were then removed by successive rounds of extraction using phenol and chloroform. Afterwards, genomic DNA was recovered by precipitation using isopropanol and then dissolved in TE buffer.

Polymerase Chain Reaction (PCR) and sequencing

The PCR amplification was carried out on a PCR system 2700 thermocycler (Applied Biosystems, Foster City, CA). The 50 μl reaction mixture contained 100 ng of fungal DNA as template, 0.2 mM of primers, 2 units of Taq DNA polymerase and 800 mM dNTPs. The 11 primer sets listed in **Table 2** were used to carry out the citrinin biosynthetic gene amplification in this study. The reaction conditions included an initial denaturation for 5 min at 96 ^{o}C, which was followed by 30 cycles for 1 min at 96 ^{o}C, 1 min at 50 ^{o}C, and 2 min at 72 ^{o}C with a final extension of 10 min at 72 ^{o}C. The PCR products were resolved and recovered on 1.2% agarose gel. Cycle sequencing reactions were then carried out using the BigDye, V3.0 kit, while DNA sequencing was performed using an ABI Prism 3730 Sequencer (Applied Biosystems, Foster City, CA).

Southern hybridization

For Southern hybridizations, genomic DNA (10 μg per lane) was digested with *Hin*dIII restriction enzymes and separated through 1.2% agarose gels through electrophoresis. Southern hybridization analysis was performed using the DIG system (Roche Diagnostics, Mannheim, Germany). The probes of the citrinin biosynthesis genes were labeled through PCR amplification from the genomic DNA of *M. purpureus* BCRC33325 (IFO30873) using the PCR DIG probe synthesis kit

(Roche Diagnostics, Mannheim, Germany). The primer sets of citrinin biosynthesis genes were *pksCT*-F8: AACTGGTCTCTTCCCCAAGC, *pksCT*-R12: ATACTGCATCGCAAACAGCG, *ctnA*-F12: TCGTTATCTAGGCTGGGCCA, *ctnA*-R1: CGTCTGGTGCAGTTAATGCG, *orf3*-F5: CGTGCACCTCTACAGGGTTC, and *orf3*-R13: GCCGCCCCATTGAAGAATAC.

Phylogenetic analysis

Sequences analyses of the amplified DNA were performed using the VectorNTI 9.0 software (InforMax, Frederick, MD). The accession numbers used for the partial beta-tubulin genes were as follows: *Monascus* species—DQ299886 to DQ299896, AY498587 to AY498589, AY498596, AY498598, AY498601, AY498602 and AY498604; *Aspergillus flavus* (M38265), *Aspergillus parasiticus* (L49386), and *Aspergillus fumigatus* (AY048754). The accession numbers used for the polyketide synthase genes were as follows: *Saccharopolyspora erythraea* DEBS (X56107 and X62569), *Aspergillus terreus lovF* (AF141925), *Aspergillus terreus lovB* (AF151722), *Penicillium citrinum mlcA*, *mlcB* (AB072893), *Phoma* sp. SQTKS (AY217789), *Cochliobolus heterostrophus pks1* (U68040), *Gibberella moniliformis* FUM1 (AF155773), *Monascus purpureus pksCT* (AB167465), *Emericella nidulans wA* (X65866), *Emericella nidulans stcA* (AAC49191), *Aspergillus parasiticus aflC* (AY371490), *Penicillium patulum* 6-MSAS (X55776), *Aspergillus parasiticus pksL2* (U52151), and *Aspergillus terreus pksM* (U31329). The phylogenetic tree was constructed using the neighbor-joining method (Saitou and Nei, 1987) via the MEGA 3.1 software with 1000 bootstrap replicates.

Measurement of citrinin

The aliquots of the *M. pilosus* culture were cleared of cells, and were then

filtered using a 0.2 mm filter. The supernatants were analyzed using HPLC performed on a Waters system (Waters, Milford, MA) fitted with a μBondapak C_{18} (10 μm) column (Waters, Milford, MA). The HPLC parameters were as follows: solvent gradient with phosphoric acid/acetonitrile/2-propanol, 60:35:5 to 25:70:5 (v:v:v) in 10 min; flow rate, 0.8 mL min^{-1}. Fluorescence detection was performed using the 2475 Multi λ Fluorescence (Waters, Milford, MA) set at 330 nm excitation wavelength and 500 nm emission wavelength. The standard citrinin compound (Sigma, St. Louis, USA) was used to confirm the HPLC analysis.

Measurement of monacolin K

The aliquots of *M. pilosus* culture were cleared of cells and filtered through a 0.2 mm filter. The supernatants were analyzed by high performance liquid chromatography (HPLC) performed on a Waters system (Waters, Milford, MA) fitted with a reverse-phase C_{18} column (LichroCART 250-4, Rp-18e, 5 μm). The HPLC parameters were as follows: solvent A, 0.1% phosphorus acid in water; solvent B, acetonitrile; 35 % A and 65 % B in 30 min; flow rate, 1.5 mL min^{-1}; and detection by UV spectroscopy (Waters 600 pump and 996 pgotodiode array detector). A standard monacolin K compound (Sigma, St. Louis, USA) was used to confirm the analysis by HPLC.

Results

PCR and Southern hybridization analysis of citrinin biosynthetic genes

Studies on fungal polyketide biosynthetic genes have indicated that metabolites are largely synthesized by the iterative multifunctional polyketide synthase systems

(Pfeifer and Khosla, 2001). Each PKS minimally carries keto-synthase (KS), acyl-transferase (AT), and acyl carrier protein (ACP) domains to catalyze different modifications. To search for the genes related to citrinin biosynthesis, 11 primer sets were designed according to the citrinin gene cluster in *M. purpureus* BCRC33325 (IFO30873) (Shimizu et al., 2005; Shimizu et al., 2007) (**Table 2** and **Figure 1**). Eighteen strains of the *Monascus* species were used to amplify the genomic DNA of the citrinin biosynthetic genes (**Table 1**). The PCR results showed that the five primer sets of the *pksCT* gene were amplified in *M. purpureus* BCRC31542, BCRC31541, BCRC33325, and BCRC31615; in *M. kaoliang* BCRC31506; and in *M. sanguineus* BCRC33446. In contrast, none of the five primer sets amplified the *pksCT* gene from *M. pilosus*, *M. ruber*, *M. barkeri*, *M. floridanus*, *M. lunisporas*, and *M. pallens*. The PCR analyses of *ctnA* encoding transcriptional activator and *orf3* encoding oxygenase with six primer sets were conducted to verify further the citrinin biosynthetic gene cluster. It was found that all strains of *M. purpureus* and *M. kaoliang* included both *ctnA* and *orf3* genes, but the two genes were absent in the strain of *M. sanguineus* and the other *Monascus* species. Although the PCR condition was adopted in a low annealing temperature (50 ^{o}C), *M. purpureus* and *M. kaoliang* can only yield the PCR products typical of *pksCT*, *ctnA*, and *orf3* genes, whereas *M. sanguineus* lacked the PCR products for *ctnA* and *orf3* genes.

To confirm the citrinin biosynthetic genes distribution in the *Monascus* species, the genomic DNA was digested by *Hin*dIII for Southern hybridization. The results revealed that the presence of *pksCT*, *ctnA* and *orf3* was consistent with the PCR results. Nevertheless, instead of a 3-kb fragment corresponding to the *pksCT* gene in *M. purpureus* and *M. kaoliang*, a 2-kb *Hin*dIII fragment was present in the *M. sanguineus*. Meanwhile, the 6.1-kb fragments corresponding to the *ctnA* and *orf3* were present in the Southern blot results of *M. purpureus* and *M. kaoliang*. It

generally agreed with the size described in the citrinin biosynthetic gene cluster from *M. purpureus* BCRC33325 (IFO30873) (**Figure 2**) (Shimizu et al., 2005). In addition, the *ctnA* and *orf3* genes were consistently absent in *M. sanguineus* and other *Monascus* species.

Sequence and phylogenetic analysis of *pksCT*, *ctnA*, *orf3*, and β-tubulin genes

The candidate PCR products of *pksCT*, *ctnA* and *orf3* were further sequenced and analyzed. The result of the 1-kb amplified DNA showed that the AT domain of *pksCT* shared a high similarity among *M. purpureus* BCRC31542, BCRC31541, BCRC33325, BCRC31615, *M. kaoliang* BCRC31506 and *M. sanguineus* BCRC33446. Both *M. purpureus* and *M. kaoliang* shared 100 % identity, while *M. sanguineus* shared 97% identity with the other species. The sequences from *M. purpureus*, *M. kaoliang*, and *M. sanguineus* all contained the conservation of deduced amino acid sequences, GHSXG, which was the typical active site of the AT domain. The phylogeny was constructed according to the conserved AT domain (**Figure 3**). The *pksCT* from *M. purpureus*, *M. kaoliang* and *M. sanguineus* belonged to the structural type as non-reduced polyketide. *M. sanguineus* was placed in a branch that was separate from *M. purpureus* and *M. kaoliang*. The KS domain of *pksCT* was further sequenced. The sequences from *M. purpureus* and *M. kaoliang* contained the conservation of deduced amino acid sequences, DXACXS, which was the typical active site of the KS domain. However, the sequence revealed a gap in the upstream of the KS domain of *M. sanguineus* (**Figure 4**); the out-of-frame *pksCT* of which may result in the nonfunctional polyketide synthase. The active site of the KS domain, DXPCXS, was also different from the typical conservation. Obviously, such *pksCT* was divergent with *M. purpureus* and *M. kaoliang.*

The amplified DNA results also showed that the *ctnA* gene shared 100 %

similarity among *M. purpureus* BCRC31542, BCRC31541, BCRC33325, and BCRC31615, as well as *M. kaoliang* BCRC31506. The *ctnA* gene-encoded transcription factor is suggested to be a positive regulatory protein involved in the citrinin biosynthesis in *M. purpureus* BCRC33325 (IFO30873) (Shimizu et al., 2007). The arrangement of the cysteine-rich nucleotide-binding domain indicated that the consensus sequence $CX_2CX_6CX_6CX_2CX_6C$ represented a Zn_2Cys_6 type zinc finger (Shimizu et al., 2007).

A phylogenetic characterization using the partial β-tubulin gene as a molecular differentiation marker was likewise conducted to further explore the evolutionary history of the various *Monascus* species. The result of the phylogenetic analysis of the partial β-tubulin gene placed both *M. purpureus* and *M. kaoliang* in the same clade (**Figure 5**). On the other hand, *M. ruber*, *M. pilosus* and *M. barkeri* were placed in the same clade. In addition, *M. sanguineus* was placed in a branch separate from *M. purpureus* and *M. kaoliang.*

PCR and sequence analyses of *mokA* and *mokE*

The PCR and sequence results showed that the AT domain of 1-kb PCR product of the *mokA* gene was amplified in *M. pilosus* BCRC31502, and BCRC38072; in *M. ruber* BCRC31523, BCRC31533, BCRC31534, BCRC31535, BCRC33314, and BCRC33323; and in *M. barkeri* BCRC33309. In contrast, none of the primer set amplified the *mokA* gene from *M. purpureus*, *M. kaoliang*, *M. sanguineus*, *M. floridanus*, *M. lunisporas*, and *M. pallens*. The PCR and sequence analyses of *mokE* encoding dehydrogenase were conducted to verify further the monacolin K biosynthetic gene cluster. It was found that all strains of *M. pilosus*, *M. ruber* and *M. barkeri* included *mokE* genes, but were absent in the other *Monascus* species (**Table 3**). This finding is in agreement with the result of the *mokA* gene.

Analysis of citrinin and monacolin K production

In previous studies, the citrinin production in *M. purpureus* BCRC33325 (IFO30873) has been shown to progress with the duration of cultivation (Shimizu et al., 2005; Shimizu et al., 2007). In this study, the amounts of citrinin produced from *Monascus* spp. were determined by HPLC after the 14 days of cultivation. The results indicated that citrinin was only detected in *M. purpureus* BCRC31542, BCRC31541, BCRC33325, and BCRC31615, as well as in *M. kaoliang* BCRC31506 (**Table 4**). In contrast, the other *Monascus* species did not produce any citrinin. The amounts of monacolin K produced from *Monascus* spp. were also detected after the 14 days cultivation. The results indicated that monacolin K was found in *M. pilosus* BCRC38072, *M. ruber* BCRC31533, BCRC31523, BCRC31534, BCRC31535 and BCRC33323.

Discussion

Citrinin is a known hepato-nephrotoxin found in the *Aspergillus*, *Penicillium* and *Monascus* species, and has been identified as a contaminant in several foods (Sweeney and Dobson, 1998; Malmstrøm et al., 2000). Citrinin accumulation in the mitochondria induces apoptosis at the cellular level (Yu et al., 2006; Chan, 2007). *Monascus*-related products have been used as herbal medicine due to the fact that its monacolin K content has shown the cholesterol synthesis inhibitor capability (Endo, 1979). To avoid the negative factor of citrinin, it is important to identify the non-citrinin producing *Monascus* strains or generally eliminate the production of citrinin in *Monascus*. Recently, the citrinin biosynthetic gene cluster has revealed

that citrinin is synthesized by polyketide synthase in *M. purpureus* BCRC33325 (IFO30873) (Shimizu et al., 2005). Moreover, the transcription factor, Zn(II)2Cys6 binuclear DNA binding protein, is also involved in the regulation of citrinin biosynthesis (Shimizu et al., 2007). In this study, the distribution of the citrinin biosynthetic genes (**Figure 1**) and the production of citrinin were detected in various *Monascus* species.

Interestingly, the PCR and Southern hybridization results demonstrated that only *M. purpureus*, *M. kaoliang*, and *M. sanguineus* contained the *pksCT* gene encoding polyketide synthase of citrinin (**Figure 2**). The sequences of the AT domain of these amplified *pksCT* genes verified that they were highly homologous with the known citrinin gene cluster in *M. purpureus* BCRC33325 (IFO30873) (**Figure 3**). However, a gap was found in the upstream of the *pksCT* KS domain of *M. sanguineus* BCRC33446. This missing base could cause a frame shift, and could not be translated as a functional polyketide synthase for citrinin biosynthesis (**Figure 4**). In addition, *M. sanguineus* BCRC33446 did not contain the *ctnA* gene encoding Zn(II)2Cys6 binuclear DNA binding protein and *orf3* encoding oxygenase (**Figure 2**). The strain may have had a large deletion of the citrinin biosynthesis gene cluster possibly initiated by a chromosomal breakage, which is similar to the *Aspergillus oryzae* RIB with a deletion at the aflatoxin biosynthesis gene cluster (Lee et al., 2006). The sequences of the 5' *ctnA* gene from *M. purpureus* and *M. kaoliang* were also verified including the consensus sequence, $CX_2CX_6CX_6CX_2CX_6C$, representing a Zn_2Cys_6 type zinc finger. The citrinin productivity phenotype was detected in *M. purpureus* and *M. kaoliang*, whereas neither *M. sanguineus* BCRC33446 nor the other *Monascus* species produced citrinin. Apparently, these results suggested that the citrinin gene cluster was highly conserved within *M. purpureus* and *M. kaoliang*, while the *pksCT* gene was shown to have high homology in *M. purpureus*, *M.*

kaoliang and *M. sanguineus*. The citrinin production results were consistent with the distribution of functional citrinin gene cluster, which was only detected in the strains of *M. purpureus* and *M. kaoliang* (**Table 1**).

Blanc et al. have initially proposed that the structure of Monascidin A is identical to the citrinin from both the *M. ruber* ATCC96218 and *M. purpureus* CBS109.07 strains (Blanc et al., 1995a). The citrinin production of *M. ruber* ATCC96218 is further investigated for biosynthetic pathway, effects of medium-chain fatty acids, and for further understanding of the improvement of the red pigment/citrinin production ratio (Hajjaj et al., 1999a; Hajjaj et al., 1999b; Hajjaj et al., 2000). Although previous studies have shown that *M. ruber* ATCC96218 can produce citrinin, this strain has unfortunately been misidentified and corrected to *M. purpureus* (Park and Jong, 2003; Park et al., 2004). Park et al. (2003; 2004) have indicated that the D1/D2 regions of the large subunit (LSU) rRNA gene and that the β-tubulin gene could be adopted to examine the phylogenetic relationship between the *Monascus* species. Their findings are in agreement with our result that *M. ruber* and *M. purpureus* can be distinguished into two series (**Figure 5**). Moreover, the *M. ruber* series (*M. ruber*, *M. pilosus* and *M. barkeri*) and the *M. purpureus* series (*M. purpureus* and *M. kaoliang*) can be grouped by determining the presence or absence of MRT non-LTR retrotransposon in the hybridization pattern according to the phylogenetic study established with the β-tubulin gene (Chen et al., 2007). Hence, *M. ruber* ATCC96218 is identical to the *M. purpureus* series based on the D1/D2 region of the LSU rRNA gene (Park et al., 2003). In addition, the strains of *M. pilosus* ATCC62949 and *M. purpureus* BCRC31523 (ATCC16378) are renamed as *M. purpureus* ATCC62949 and *M. ruber* BCRC31523 (ATCC16378), respectively (Park et al., 2003). Citrinin is only produced in *Penicillium citrinum* among the detection of *P. citrinum, P. corylophilum, P. steckii, P. sizovae* and *P. sumatrense* with 79

isolates (Malmstrøm et al., 2000). It has also been reported in *P. expansum* and *P. verrucosum* (Sweeney and Dobson, 1998). These results shown that citrinin was produced in certain species of *Penicillium*. This is quite similar to our findings that only the species of *M. purpureus* and *M. kaoliang* can produce citrinin.

Wang et al. have proposed that citrinin is detectable in each *Monascus* species under a low HPLC flow rate condition (0.15 mL min^{-1}) and incubated in a YES medium (40 g $liter^{-1}$ of yeast extract and 160g $liter^{-1}$ of sucrose) (Wang et al., 2005). However, these results need to be rechecked because we did not detect citrinin in some of the species in the list included in this paper. Moreover, their incubation and analysis condition might cause false positive results. It is possible, but unlikely, that another citrinin biosynthesis gene was divergent and present in the *M. ruber* series.

In conclusion, the citrinin production in *Monascus* was consistent with the presence of the functional citrinin biosynthetic genes found only in the *Monascus purpureus* series strains. Although *M. sanguineus* BCRC 33446 carried the *pksCT* gene, citrinin productivity was not detected due to a frame shift of *pksCT* and the absence of some other citrinin biosynthetic genes (**Figure 2** and **Figure 4**). The *M. ruber*, *M. pilosus* and *M. barkeri* strains were grouped together based on the classification of the β-tubulin gene. They were found to be neither *pksCT* , *ctnA*, and *orf3* genes nor citrinin productivity. Therefore, citrinin biosynthetic genes and citrinin productivity can also be used for phylogenetic study and in the exploration of gene evolution for secondary metabolism.

References

(1) Babitha, S.; Soccol, C. R.; Pandey, A. Solid-state fermentation for the production of *Monascus* pigments from jackfruit seed. *Bioresour. Technol.* **2007**, *98*, 1554-1560.

(2) Bingle, L. E. H.; Simpson, T. J.; Lazarus, C. M. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet. Biol.* **1999**, *26*, 209-223.

(3) Blanc, P. J.; Laussac, J. P.; Le Bars, J.; Le Bars., P.; Loret, M. O.; Pareilleux, A.; Prome, D.; Prome, J. C.; Santerre, A. L.; Goma, G. Characterization of monascidin A from *Monascus* as citrinin. *Int. J. Food Microbiol.* **1995a**, *27*, 201-213.

(4) Blanc, P. J.; Loret, M. O.; Goma, G. Production of citrinin by various species of *Monascus*. *Biotechnol. Lett.* **1995b**, *17*, 291-294.

(5) Chan, W. H. Citrinin induces apoptosis via a mitochondria-dependent pathway and inhibition of survival signals in embryonic stem cells, and causes developmental injury in blastocysts. *Biochem. J.* **2007**, *404*, 317-326.

(6) Chen, Y.-P.; Tseng, C.-P.; Liaw, L.-L.; Wang, C.-L.; Yuan, G.-F. Characterization of MRT, a New Non-LTR Retrotransposon in *Monascus* spp. *Bot. Bull. Acad. Sin.* **2007**, *48*, 377-385.

(7) Endo, A. Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* (Tokyo) **1979**, *32*, 852-854.

(8) Endo, A.; Komagata, D.; Shimada, H. Monacolin M, a new inhibitor of cholesterol biosynthesis. *J. Antibiot.* (Tokyo) **1986**, *39*, 1670-1673.

(9) Hajjaj, H.; Klaebe, A.; Loret, M. O.; Goma, G.; Blanc, P. J.; Francois, J. Biosynthetic pathway of citrinin in the filamentous fungus *Monascus ruber* as revealed by ^{13}C nuclear magnetic resonance. *Appl. Environ. Microbiol.* **1999a**, *65*,

311-314.

(10) Hajjaj, H.; Blanc, P. J.; Groussac, E.; Goma, G.; Uribelarrea, J. L.; Loubiere, P. Improvement of red pigment/citrinin production ratio as a function of environmental conditions by *Monascus ruber*. *Biotechnol. Bioeng.* **1999b**, *64*, 497-501.

(11) Hajjaj, H.; Klaébé, A.; Goma, G.; Blanc, P. J.; Barbier, E.; François, J. Medium-chain fatty acids affect citrinin production in the filamentous fungus *Monascus ruber*. *Appl. Environ. Microbiol.* **2000**, *66*, 1120-1125.

(12) Hendrickson, L.; Davis, C. R.; Roach, C.; Nguyen, D. K.; Aldrich, T.; McAda, P. C.; Reeves, C. D. Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chem. Biol.* **1999**, *6*, 429-439.

(13) Kennedy, J.; Auclair, K.; Kendrew, S. G.; Park, C.; Vederas, J. C.; Hutchinson, C. R. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science.* **1999**, *284*, 1368-72.

(14) Lee, Y. H.; Tominaga, M.; Hayashi, R.; Sakamoto, K.; Yamada, O.; Akita, O. *Aspergillus oryzae* strains with a large deletion of the aflatoxin biosynthetic homologous gene cluster differentiated by chromosomal breakage. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 339-345.

(15) Malmstrøm, J.; Christophersen, C.; Frisvad, J. C. Secondary metabolites characteristic of *Penicillium citrinum*, *Penicillium steckii* and related species. *Phytochemistry.* **2000**, *54*, 301-309.

(16) Park, H. G.; Jong, S.-C. Molecular characterization of *Monascus* strains based on the D1/D2 regions of LSU rRNA genes. *Mycoscience.* **2003**, *44*, 25-32.

(17) Park, H. G.; Stamenova, E. K.; Jong, S.-C. Phylogenetic relationships of *Monascus* species inferred from the ITS and the partial β-tubulin gene. *Bot. Bull.*

Acad. Sin. **2004**, *45*, 325-330.

(18) Pfeifer, B. A.; Khosla, C. Biosynthesis of polyketides in heterologous hosts. Microbiol. *Mol. Biol. Rev.* **2001**, *65*, 106-118.

(19) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406-425.

(20) Shimizu, T.; Kinoshita, H.; Ishihara, S.; Sakai, K.; Nagai, S.; Nihira, T. Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2005**, *71*, 3453-3457.

(21) Shimizu, T.; Kinoshita, H.; Nihira, T. Identification and in vivo functional analysis by gene disruption of *ctnA*, an activator gene involved in citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* **2007**, *73*, 5097-5103.

(22) Sweeney, M. J.; Dobson, A. D. Mycotoxin production by *Aspergillus*, *Fusarium* and *Penicillium* species. *Int. J. Food Microbiol.* **1998**, *43*, 141-158.

(23) Tobert, J. A. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* **2003**, *2*, 517-526.

(24) Wang, Y.-Z.; Ju, X.-L.; Zhou, Y.-G. The variability of citrinin production in *Monascus* type cultures. *Food Microbiology.* **2005**, *22*, 145-148.

(25) Yu, F. Y.; Liao, Y. C.; Chang, C. H.; Liu, B. H. Citrinin induces apoptosis in HL-60 cells via activation of the mitochondrial pathway. *Toxicol. Lett.* **2006**, *161*, 143-151.

Table 1. Strains used in this work for determination of citrinin.

Strain	Species[a]	polyketide synthase (*pksCT*)[b]	Activator (*ctnA*)	Oxygenase (*orf3*)	Source	Citrinin production
1. BCRC 31502 (ATCC 16363)	*Monascus pilosus*, Type	—	—	—	Japan	—
2. BCRC 38072	*Monascus pilosus*	—	—	—	Taiwan	—
3. BCRC 31533 (ATCC 16246)	*Monascus ruber*, Type	—	—	—	Nigeria	—
4. BCRC 31523 (ATCC 16378)	*Monascus ruber*	—	—	—	Taiwan	—
5. BCRC 31534 (ATCC 16366)	*Monascus ruber*	—	—	—	Switzerland	—
6. BCRC 31535 (ATCC 18199)	*Monascus ruber*	—	—	—	Canada	—
7. BCRC 33314 (ATCC 16371)	*Monascus ruber*	—	—	—	Taiwan	—
8. BCRC 33323 (ATCC 18199)	*Monascus ruber*	—	—	—	Canada	—
9. BCRC 31542 (ATCC 16365)	*Monascus purpureus*, Type	+	+	+	Java	+
10. BCRC 31541 (ATCC 16379)	*Monascus purpureus*	+	+	+	Taiwan	+
11. BCRC 33325 (IFO 30873)	*Monascus purpureus*	+	+	+		+
12. BCRC 31615 (DSM 1379)	*Monascus purpureus*	+	+	+		+

13. BCRC 31506 (CBS 302.78)	*Monascus kaoliang*, Type	+	+	+	Taiwan	+
14. BCRC 33446 (ATCC 200613)	*Monascus sanguineus*, Type	+	—	—	Iraq	—
15. BCRC 33309 (ATCC 16966)	*Monascus barkeri*	—	—	—	Japan	—
16. BCRC 33310 (IMI 282587)	*Monascus floridanus*, Type	—	—	—	USA	—
17. BCRC 33640 (ATCC 204397)	*Monascus lunisporas*, Type	—	—	—	Japan	—
18. BCRC 33641 (ATCC 200612)	*Monascus pallens*, Type	—	—	—	Iraq	—

[a] "Type" indicates type strain.

[b] +, positive; —, negative.

Table 2. Primers used to amplify citrinin related genes fragments.

Primers[a]	Sequence (5' → 3')	Position[b]
F1	AACGGACAGGAAGAGCGTGC (20-mer)	68 to 87 (*ctnA*)
F2	ACGAGTGTCAGTTCGGCTCC (20-mer)	483 to 502 (*ctnA*)
F3	TCGGAAGCGATCATGGACGT (20-mer)	1277 to 1296 (*ctnA*)
F4	CTCCTTTCCGCGCAATTCCA (20-mer)	760 to 779 (*ctnA*)
F5	CGTGCACCTCTACAGGGTTC (20-mer)	113 to 132 (*orf3*)
F6	CTACCAGGCCATGCTGAAGC (20-mer)	447 to 466 (*orf3*)
F7	GAGTCCCCGAGAAATGGCAT (20-mer)	1445 to 1464 (*pksCT*)
F8	AACTGGTCTCTTCCCCAAGC (20-mer)	2661 to 2680 (*pksCT*)
F9	TTAACCGTCTCCTGTCCGGC (20-mer)	4012 to 4031 (*pksCT*)
F10	TGCCTATCACGTCAACGGCA (20-mer)	5348 to 5367 (*pksCT*)
F11	ACGTGGACCATGCCGAGAAC (20-mer)	3313 to 3332 (*pksCT*)
R1	CGTCTGGTGCAGTTAATGCG (20-mer)	958 to 977 (*ctnA*)
R2	GGTATGGCATCGGTGGTGTG (20-mer)	1568 to 1587 (*ctnA*)
R3	GAAACGGGGGAGTGGATTGG (20-mer)	700 to 719 (*orf3*)
R4	GAGGATCGGATGCGGCATTT (20-mer)	1843 to 1862 (*ctnA*)
R5	TCTTCGATGGCAACCTGGAC (20-mer)	854 to 873 (*orf3*)
R6	CTGCCATCTTCCAAGCCCAA (20-mer)	239 to 258 (*orf4*)
R7	AACAACTACGGTGCTTCCGG (20-mer)	2427 to 2446 (*pksCT*)
R8	CGGGCTCTGGGTACATCAAA (20-mer)	3654 to 3673 (*pksCT*)
R9	CGGTCTTGAACCTGACGAGG (20-mer)	5090 to 5109 (*pksCT*)
R10	GAAGGTACTCGGCCAGAAGC (20-mer)	6486 to 6505 (*pksCT*)
R11	CAATCACATTCCAAGCGGCG (20-mer)	4303 to 4322 (*pksCT*)
F12	TCGTTATCTAGGCTGGGCCA (20-mer)	678 to 697 (*ctnA*)

R12	CGCTGTTTGCGATGCAGTAT (20-mer)	2977 to 2996 (*pksCT*)
R13	GCCGCCCCATTGAAGAATAC (20-mer)	428 to 447 (*orf3*)

[a] F, forward primer; R, reverse primer.

[b] The sites corresponding to those of the *M. purpureus* BCRC33325 (IFO30873) citrinin gene cluster (GenBank accession no. AB243687).

Table 3. Strains used in this work for determination of monacolin K.

Strain	Species[a]	polyketide synthase (*mokA*)[b]	dehydrogenase (*mokE*)	Source	Monacolin K production
1. BCRC 31502 (ATCC 16363)	*Monascus pilosus*, Type	+	+	Japan	—
2. BCRC 38072	*Monascus pilosus*	+	+	Taiwan	+
3. BCRC 31533 (ATCC 16246)	*Monascus ruber*, Type	+	+	Nigeria	+
4. BCRC 31523 (ATCC 16378)	*Monascus ruber*	+	+	Taiwan	+
5. BCRC 31534 (ATCC 16366)	*Monascus ruber*	+	+	Switzerland	+
6. BCRC 31535 (ATCC 18199)	*Monascus ruber*	+	+	Canada	+
7. BCRC 33314 (ATCC 16371)	*Monascus ruber*	+	+	Taiwan	—
8. BCRC 33323 (ATCC 18199)	*Monascus ruber*	+	+	Canada	+
9. BCRC 31542 (ATCC 16365)	*Monascus purpureus*, Type	—	—	Java	—
10. BCRC 31541 (ATCC 16379)	*Monascus purpureus*	—	—	Taiwan	—
11. BCRC 33325 (IFO 30873)	*Monascus purpureus*	—	—		—
12. BCRC 31615 (DSM 1379)	*Monascus purpureus*	—	—		—

13. BCRC 31506 (CBS 302.78)	*Monascus kaoliang*, Type	—	—	Taiwan	—
14. BCRC 33446 (ATCC 200613)	*Monascus sanguineus*, Type	—	—	Iraq	—
15. BCRC 33309 (ATCC 16966)	*Monascus barkeri*	+	+	Japan	—
16. BCRC 33310 (IMI 282587)	*Monascus floridanus*, Type	—	—	USA	—
17. BCRC 33640 (ATCC 204397)	*Monascus lunisporas*, Type	—	—	Japan	—
18. BCRC 33641 (ATCC 200612)	*Monascus pallens*, Type	—	—	Iraq	—

[a] "Type" indicates type strain.

[b] +, positive; —, negative.

Table 4. The concentration of citrinin produced by *Monascus* species.

Strains[a]	*M. purpureus*				*M. kaoliang*
	BCRC31542	BCRC31541	BCRC33325	BCRC31615	BCRC31506
Citrinin (mg/g mycelia)[b]	0.33 ± 0.14	15.35 ± 6.17	6.23 ± 1.74	6.56 ± 2.82	2.71 ± 0.28

[a] *Monascus* species were harvested after the cultivation of 14 days.

[b] Citrinin was detected by HPLC with the fluorescence detector under culture condition described in material and methods.

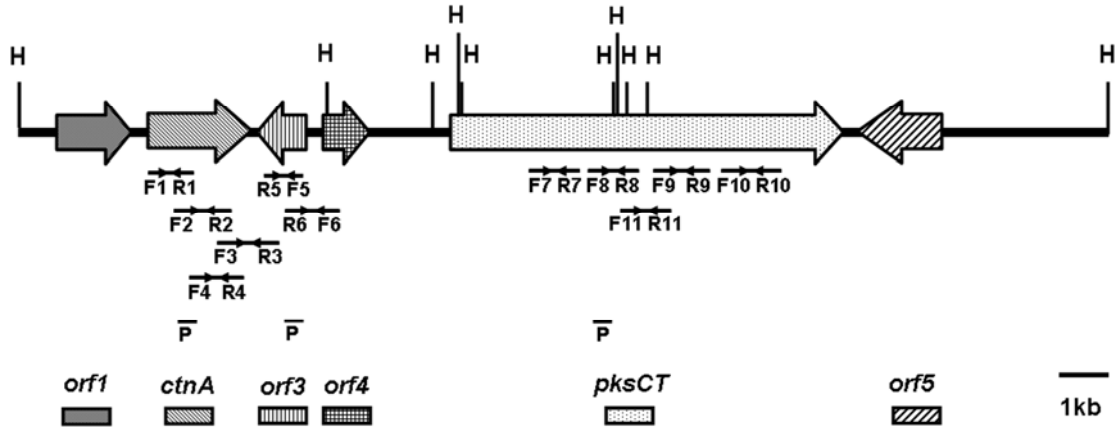


Figure 1. The citrinin gene cluster of *M. purpureus* BCRC33325 (IFO30873) was obtained from the GenBank database using accession number AB243687. The citrinin gene cluster includes Dehydrogenase (*orf1*), Transcriptional activator (*ctnA*), Oxygenase (*orf3*), Oxidoreductase (*orf4*), Polyketide synthase (*pksCT*), and Transporter (*orf5*). The abbreviations F and R indicate the forward and reverse primers, respectively, of *ctnA*, *orf3*, and *pksCT* for the PCR analysis; the abbreviation P indicates the probes of *ctnA*, *orf3*, and *pksCT* for the Southern hybridization analysis.

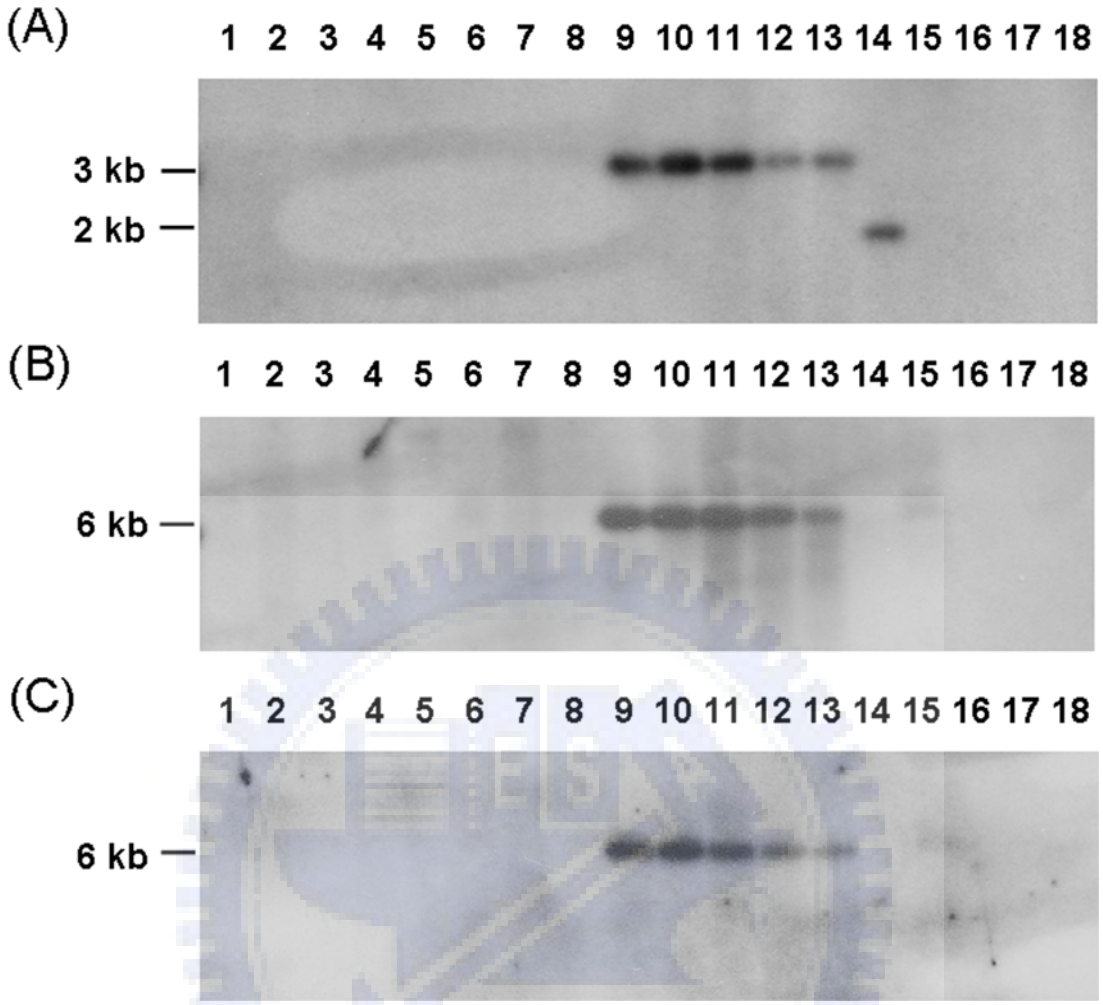


Figure 2. The Southern hybridization analyses of the citrinin-related genes. Chromosome DNAs extracted from the *Monascus* species were digested by *Hin*dIII, separated on gel electrophoresis, and hybridized with the *ctnA*, *orf3*, and *pksCT* probes. *Monascus* species— lane 1: *M. pilosus* BCRC31502; lane 2: *M. pilosus* BCRC38072; lane 3: *M. ruber* BCRC31533; lane 4: *M. ruber* BCRC31523; lane 5: *M. ruber* BCRC31534; lane 6: *M. ruber* BCRC31535; lane 7: *M. ruber* BCRC33314; lane 8: *M. ruber* BCRC33323; lane 9: *M. purpureus* BCRC31542; lane 10: *M. purpureus* BCRC31541; lane 11: *M. purpureus* BCRC33325; lane 12: *M. purpureus* BCRC31615; lane 13: *M. kaoliang* BCRC31506; lane 14: *M. sanguineus* BCRC33446; lane 15: *M. barkeri* BCRC33309; lane 16: *M. floridanus* BCRC33310; lane 17: *M.*

lunisporas BCRC33640; and lane 18: *M. pallens* BCRC33641. (A) *pksCT* gene (B) *ctnA* gene (C) *orf3* gene.

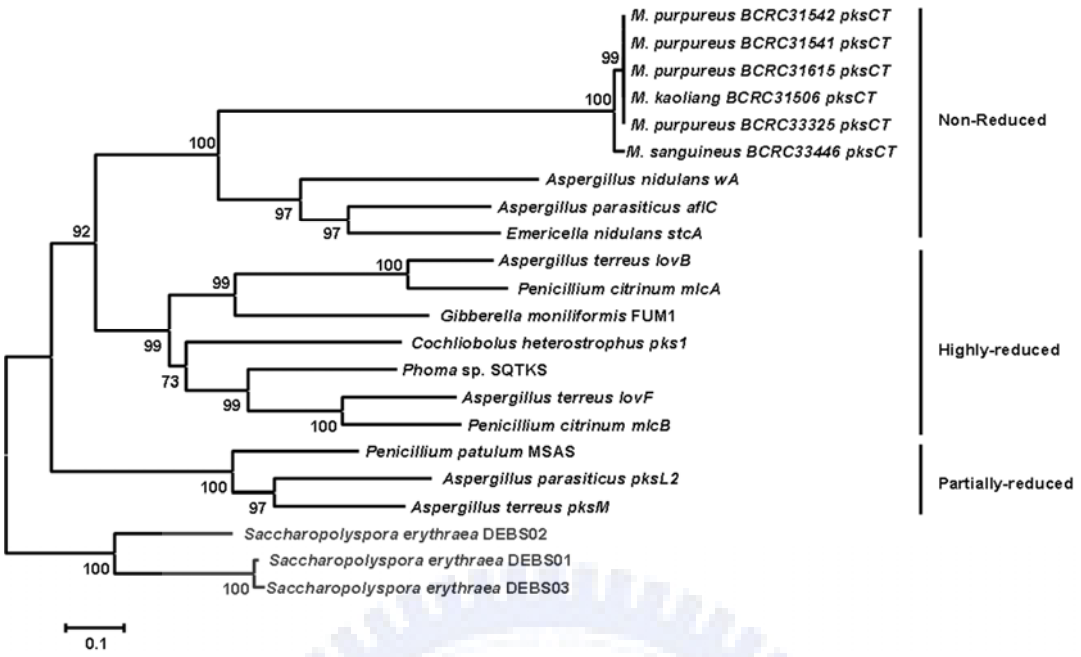


Figure 3. Phylogenetic tree of PKSs from *Monascus* and various organisms. The phylogeny of PKSs based on the conserved acyl-transferase domain with 300 amino acids was constructed and rooted using the acyl-transferase domains of *Saccharopolyspora erythraea* DEBS (X56107 and X62569). The accession numbers for the polyketide synthase genes were as follows: *Aspergillus terreus lovF* (AF141925), *Aspergillus terreus lovB* (AF151722), *Penicillium citrinum mlcA*, *mlcB* (AB072893), *Phoma* sp. SQTKS (AY217789), *Cochliobolus heterostrophus pks1* (U68040), *Gibberella moniliformis* FUM1 (AF155773), *Monascus purpureus pksCT* (AB167465), *Emericella nidulans wA* (X65866), *Emericella nidulans stcA* (AAC49191), *Aspergillus parasiticus aflC* (AY371490), *Penicillium patulum* 6-MSAS (X55776), *Aspergillus parasiticus pksL2* (U52151), and *Aspergillus terreus pksM* (U31329). Bootstrap values are shown in the nodes according to the 1000 replications. Only bootstrap values >50 % are shown.

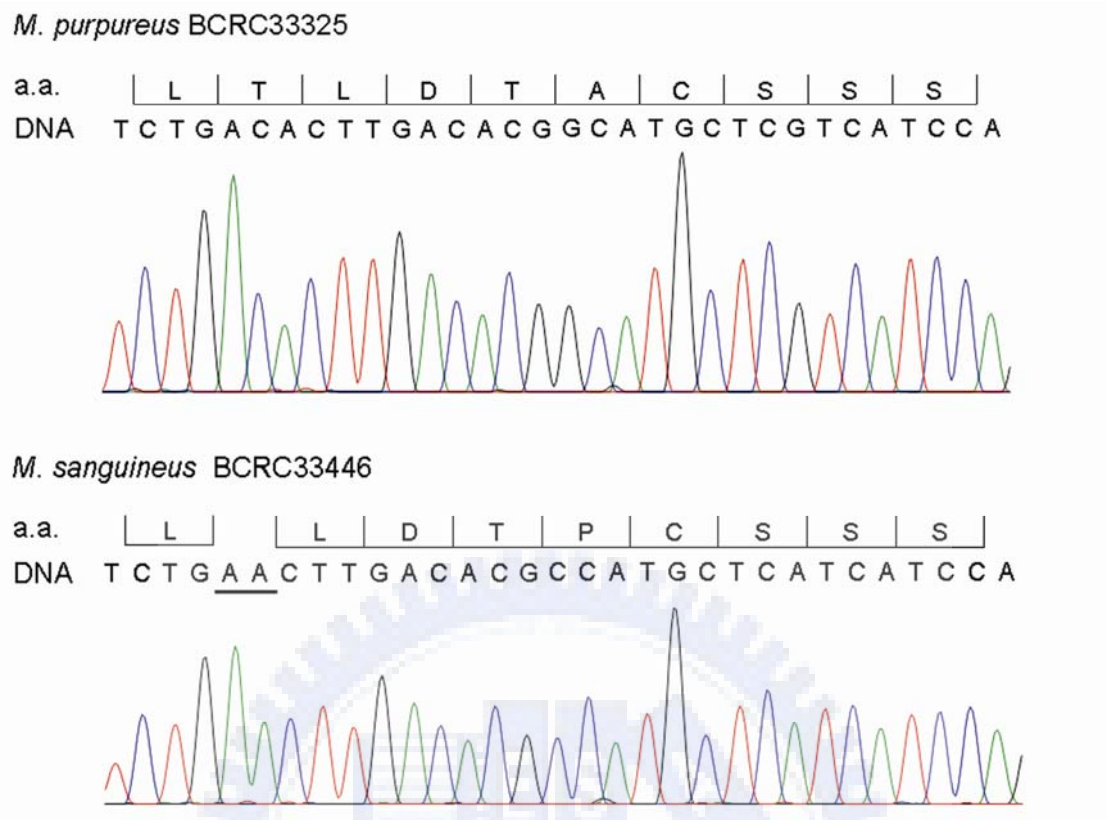


Figure 4. Sequence analyses of the *pksCT* KS domain. The conserved deduced amino acid sequences of the typical active site of the KS domain, DXACXS, were found in *M. purpureus* BCRC33325. A gap, CTGA-ACTT, was observed in the upstream of the *pksCT* KS domain of *M. sanguineus* BCRC33446. The sequence of *M. sanguineus* BCRC33446 was carried out in triplicate.

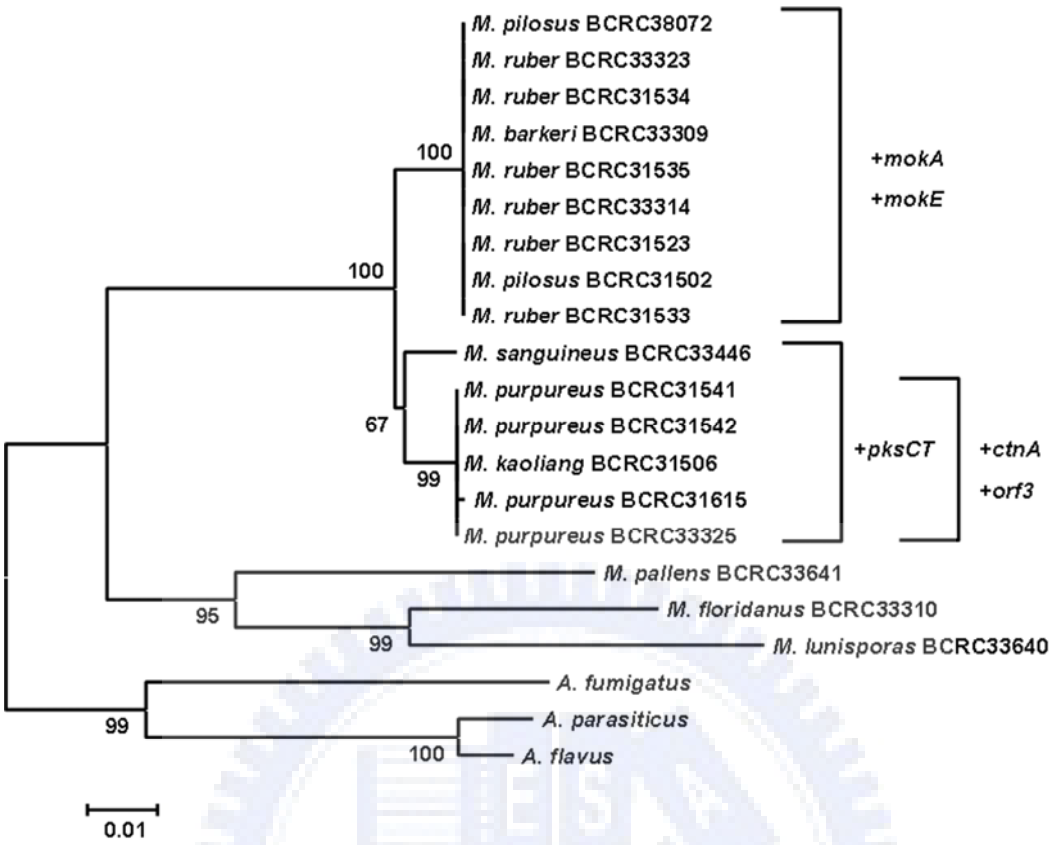


Figure 5. Phylogeny of the *Monascus* species based on the partial β-tubulin gene. The partial β-tubulin genes of the *Monascus* species were used by the following accession numbers: DQ299886 to DQ299896, AY498587 to AY498589, AY498596, AY498598, AY498601, AY498602, and AY498604. The accession numbers of the β-tubulin genes were used as outgroups of the following: *Aspergillus flavus* (M38265), *Aspergillus parasiticus* (L49386), and *Aspergillus fumigatus* (AY048754). Bootstrap values are shown in the nodes according to 1000 replications. Only bootstrap values >50 % are shown.

Chapter 5

Characterization of MRT, a New Non-LTR Retrotransposon in *Monascus* spp.

Abstract

A new non-LTR retrotransposon, named MRT, was discovered in the filamentous fungus *Monascus pilosus* BCRC38072. The entire nucleotide sequence of the MRT element was 5.5-kb long, including two open reading frames. These two ORFs showed homologies to *gag*-like and *pol*-like gene products, and an A-rich sequence at the 3' end of *pol*-like gene. ORF1 encoded a protein of 517 amino acids and contained a cysteine-rich zinc finger motif. ORF2 encoded a protein of 1181 amino acids and contained apurinic/apyrimidinic endonuclease (APE), reverse transcriptase (RT), RNaseH domains, and a CCHC motif. The phylogenetic analyses demonstrated that the MRT element was classified into the Tad1 clade. The results of Southern hybridizations showed that MRT elements were distributed within *M. pilosus*, *M. ruber*, *M. sanguineus*, and *M. barkeri*. In addition, the species of *Monascus* can be grouped by the presence or absence of MRT elements in the hybridization pattern according to phylogenetic subgroups established with the partial β-tubulin gene.

Key words: *Monascus pilosus*, Non-LTR retrotransposon, Phylogenetic analysis, Bacterial artificial chromosome

Introduction

Retrotransposons are found in most eukaryotes and in some case constitute a major part of the genome (e.g. 40-50% of the human genome). They have been divided into two subclasses based on their differences in overall structure. LTR retrotransposons closely relate to retroviruses and non-LTR retrotransposons also call LINE-like elements. All these elements use reverse transcription to propagate. Non-LTR retrotransposons have been found in many groups of eukaryotic organisms including mammals, insects, amphibians, plants and also fungi.

Five non-LTR retrotransposons have been characterized in filamentous fungi including *Tad1-1* in *Neurospora crassa* (Cambareri et al., 1994), MGR583 in *Magnaporthe grisea* (Hamer et al., 1989), CgT1 in *Colletotrichum gloeosporioides* (He et al., 1996), *marY2N* in *Tricholoma matsutake* (Murata et al., 2001), and Mars1 in *Ascobolus immerses* (Goyon et al., 1996) though the entire element is not described for the last species. These non-LTR retrotransposons usually contain two ORFs encoding *gag*-like and *pol*-like proteins. Moreover, they generally have poly (A) or A-rich region at their 3' terminus and generate truncation in 5' UTRs. Phylogenetic analysis of non-LTR retrotransposons based on the RT (reverse transcriptase) domain, the only sequence found in all elements defined 11 clades by Malik et al. (1999). More recently, Burke et al. (2002) propose an additional classification in which the various clades fall in 5 groups on the basis of both the phylogenetic relationship of their RT sequence and the nature and arrangement of their protein domains. Based on sequence, structure, and phylogenetic analyses, the non-LTRs elements from filamentous fungi are grouped in the Tad1 clade. Recently, two non-LTR retrotransposons of yeast, Zorro in *Candida albicans* (Goodwin et al., 2001) and Ylli in *Yarrowia lipolytica* (Casaregola et al., 2002), are found belonging to the L1 clade of mammalian elements.

According to the distribution of CgT1 in *C. gloeosporioides*, presence or absence of CgT1 can be used to distinguish biotypes A and B that cause different anthracnose diseases on *Stylosanthes* in Australia (He et al., 1996). Moreover, DNA fingerprints analysis of CgT1 reveals that Australian isolates of biotype B are monomorphic. In addition, the analysis of the genetic relations and evolutionary history of many species has been facilitated by repetitive DNA fingerprinting probe (Cizeron et al., 1998; Blesa et al., 2001; Daboussi and Capy, 2003).

Monascus spp. belong to ascomycetes, used in Chinese fermented foods such as anka, anka pork, rice wine etc. for thousands of years. Thirteen *Monascus* species have been reported. They are known as producers of various secondary metabolites with polyketide structures, such as monacolins, and have medical importance (Endo, 1979). In this study, we report the discovery of a new non-LTR retrotransposon named MRT (*Monascus* Retrotransposon) in *Monascus* spp.. The structural, genomic and phylogenetic analysis of the MRT elements was presented. Moreover, the distribution of MRT in *Monascus* species was also analyzed.

Materials and methods

Strains, media and growth conditions

The nineteen strains of *Monascus* listed in **Table 1** were used in this study. All strains were maintained on YM (DIFCO, Detroit, Michigan) agar for one week, and spore suspensions were obtained by washing cultured YM agar plates with distilled water. Mycelia for DNA isolation were harvested from YM broth after incubating for 8 days at 28^{o}C with constant agitation and then frozen at -80^{o}C.

BAC library construction and shotgun sequencing

The methods of Peterson et al. (2000) were used to make a BAC library of *Monascus pilosus* BCRC 38072. DNAs of eleven BAC clones were extracted for shotgun sequencing by Qiagen Large-Construct kit (Qiagen, Valencia, CA). DNA sequencing was performed with an ABI Prism 3700 Sequencer (Applied Biosystems, Foster City, CA). Phred-Phrap-Consed system developed by the Phil Green Laboratory was used to assemble DNA fragments (Gordon et al., 2001). Nucleotide and deduced amino acid sequences were used to interrogate the non-redundant database at GenBank using BlastN and BlastX. Sequence analysis was done using VectorNTI 9.0 (InforMax, Frederick, MD) software. Prediction of nucleic acid secondary structure was performed with Mfold server (http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html). The nucleotide sequences of MRT non-LTR retrotransposons found in this study have been submitted to GenBank under the accession numbers AY900582 and DQ299897 to DQ299900.

Genomic DNA preparation and Southern hybridization

Monascus genomic DNA was extracted according to the method developed by Bingle et al. (1999). Approximately 0.5 g (squeezed wet weight) of frozen mycelia was ground to a fine power under liquid nitrogen using a mortar and pestle. Protein was removed by successively rounds of extraction with phenol and chloroform. Genomic DNA was recovered by precipitation with ethanol and dissolved in TE buffer. For Southern hybridizations, genomic DNA (7.5 μg per lane) was digested with *Eco*RI and *Bam*HI restriction enzymes and separated through 1.0% agarose gels by electrophoresis. Southern hybridization analysis was performed using the DIG system (Roche Diagnostics, Mannheim, Germany). The probe of the MRT element was labeled by PCR amplification from genomic DNA of *M. pilosus* 38072 using

PCR DIG probe synthesis kit (Roche Diagnostics, Mannheim, Germany). The primer set of the MRT probe was MRT1-1F: CAGGGGGAGGCTAGGATGTA, and MRT1-1R: CACAGGTGGGTAGAGCCACAG. All other DNA manipulations were performed as described in Sambrook et al. (1989).

Phylogenetic analysis

In addition to MRT element, β-tubulin gene was chosen for phylogenetic analysis of *Monascus* spp.. The partial β-tubulin genes were amplified with the primer set, btubulinF: 5'- CAACTGGGCTAAGGGTCATT and btubulinR: 5'-GTGAACTCCATCTCGTCCATA (Wu et al., 1996; Park et al., 2004). Sequences of partial β-tubulin genes obtained from *Monascus* strains using in this study had been submitted to GenBank under the accession numbers DQ299886 to DQ299896. The other accession numbers of partial β-tubulin genes, AY498587 to AY498589, AY498596, AY498598, AY498601, AY498602 and AY498604, were obtained from the GenBank database. The phylogenetic tree was constructed by the neighbor-joining method (Saitou and Nei, 1987) using MEGA 3.1 software with 1000 bootstrap replicates.

Results and discussion

During the whole genome sequencing of *M. pilosus* BCRC38072, two repetitive sequences (mps01-1 and mps01-2) were observed in a *ca.* 160kb BAC, mps01. They were found to have a similar homology to CgT1 (He et al., 1996) which is a non-LTR retrotransposon from *Colletotrichum gloeosporioides* (mps01-1, 31% identity and mps01-2, 32% identity by BlastX). The sequence homology indicated that the repetitive sequences were non-LTR retrotransposons. The new non-LTR retrotransposon was designated as MRT. Eleven BAC clones of *M. pilosus*

BCRC38072 were sequenced covering 1.55 Mb; and six positive BACs were identified and 15 copies of the MRT element were found, in determining the relative abundance and diversity of the non-LTR retrotransposon. The entire nucleotide sequence of the MRT element was 5.5-kb long. Given that the size of the *M. pilosus* BCRC38072 genome was ~30 Mb, the number of copies of the genome was estimated to be ~290, which occupied about 5% of the *M. pilosus* BCRC38072 genome.

The translation frames of the entire set of MRT elements revealed that four of them in BACs, mps02-1, mps07-1, mps11-1, and mps13-1, contained two open reading frames (**Figure 1**), like other non-LTR retrotransposons that have been found in fungi - CgT1, Tad-1 and MGR583 (Hamer et al., 1989; Cambareri et al., 1994; He et al., 1996). Numerous stop codons were found in other copies of the MRT elements. A conceptual translation demonstrated that the first ORF in the MRT element may encode a protein of 517 amino acids. The deduced amino acid sequence of the MRT ORF1 contained one zinc finger motif that was cysteine-rich (**Figure 1, Figure 2**). The arrangement of cysteines indicated that the consensus sequence $CX_6HXCX_6CHX_2HX_6C$ represented an NF-X1-type zinc finger, based on Pfam analysis (Song et al., 1994). However, no 5' UTR was clearly identified, suggesting that all of the active element may be rare in *M. pilosus* BCRC38072. A conceptual translation demonstrated that the second ORF in the MRT element may encode a protein of 1181 amino acids. The ORF1 and ORF2 overlapped by 1 bp. The deduced amino acid sequence of the MRT ORF2 contained an apurinic/apyrimidinic endonuclease (APE) domain, a reverse transcriptase (RT) domain, an RNaseH domain, and a CCHC motif (**Figure 1, Figure 3**). The APE domain at the N-terminal of ORF2 in the MRT element was proven not to be well conserved in other organisms. The RT domain located downstream of the APE domain contained conservation of deduced amino acid sequences, YXDD, which were

a part of the active site in the RT domain (He et al., 1996). The 3' UTRs of the MRT elements were around 300 bp, and the 3' ends were well conserved. Two parts, the A-rich sequence and the stem-loop region, were present in the conserved tail (**Figure 4A**). Interestingly, the A-rich stretch represented two sequences, TAAATAATAA(CATAA)n and TAAATAATAA(A)n. Additionally, the RNA transcribed from the conserved tail of the MRT element was proposed to form a stem-loop (**Figure 4B**). This stem-loop region can be recognized by the reverse transcriptase of the non-LTR retrotransposon (Baba et al., 2004). Furthermore, most non-LTR retrotransposons have target site duplications (TSDs) of variable lengths from 4 to 49 bp (Eickbush, 1992). The sequence results also revealed that four MRT elements presented 7 to 15 bp target site duplications (**Figure 4C**).

Since reverse transcriptase (RT) is the fundamental component of the machinery that is required to synthesize DNA, it is strongly conserved and used in analyses of retrotransposon phylogeny (Flavell, 1995). In particular, the 11 conserved block sequences of the reverse transcriptase domain defined by Malik et al. (1999) are extensively used for the construction of retrotransposon phylogeny. This study analyzed the phylogenetic relationships between members of the non-LTR retrotransposons, including the four MRT elements described above, using shared reverse transcriptase domains. The phylogenetic tree was rooted using RT sequences of bacterial and fungal group II introns (Xiong and Eickbush, 1990). Eleven clades were markedly distinguished (**Figure 5A**). This result was consistent with the phylogeny constructed by Malik et al. (1999). The phylogenetic tree further suggested that the MRT element belonged to the Tad1 clade known from filamentous fungi. Moreover, the apurinic/apyrimidinic endonuclease (APE) domain is believed to cleave DNA in the reverse transcription reaction at the chromosome target site (Maita et al., 2004) and is commonly employed to construct the phylogenetic analysis

(Malik et al., 1999). Since CRE, R2 and R4 clades lacked an APE domain, other clades were applied to construct a phylogenetic tree that was rooted using apurinic/apyrimidinic endonuclease of *Drosophila* sp. (AAB19427), and *Escherichia coli* (P09030). **Figure 5B** depicted the phylogenetic tree of the non-LTR retrotransposons according to the APE domain. This result demonstrated that the close relationship between members of the Tad1 clade was consistent with the phylogeny obtained from the reverse transcriptase domain. The APE phylogeny had lower resolution than the RT phylogeny, perhaps because that APE domain was smaller and less conserved than the RT domain (Malik et al., 1999). Additionally, even though the APE and RT domains of CgT1 contained termination codons in the ORF1 and ORF2 (He et al., 1996), the phylogeny analysis of the deduced amino acid sequence of the CgT1 element also revealed that it was belonging to Tad1 clade.

In order to study the distribution of MRT elements in *Monascus* species (**Table 1**), their genomic DNA was extracted and digested by *Eco*RI and *Bam*HI. The results of Southern hybridizations showed that the MRT elements were widely distributed over *M. pilosus*, *M. ruber* and *M. barkeri* (**Figure 6**) and a large number of high-intensity bands were detected in these species. In contrast, the MRT element was absent in the species of *M. purpureus*, *M. kaoliang*, *M. floridanus*, *M. lunisporas*, and *M. pallens*. The intensity and diversity of hybridization patterns also shown in *M. sanguineus*, a newly found species of *Monascus*, were weaker than those in the species of *M. pilosus*, *M. ruber*, and *M. barkeri*. The weaker intensity of bands implied that the copy number of MRT in *M. sanguineus* was fewer than those in *M. ruber*, *M. pilosus* and *M. barkeri*. The fingerprints of the DNA hybridizations demonstrated that the band patterns were distributed between 500 bp to 10 kb. However, the species *M. pilosus* BCRC31503 was an exception that could not detect the presence of any MRT elements (**Figure 6**). A phylogenetic characterization

using the partial β-tubulin gene as a molecular differentiation marker was conducted to elucidate the evolutionary history of MRT elements among different species. According to the study of Park et al. (2004), the partial β-tubulin gene can be adopted to examine the phylogenetic relationship among the *Monascus* species without gaps in the alignment of partial β-tubulin genes. *Aspergillus flavus* (M38265), *Aspergillus parasiticus* (L49386), *Aspergillus fumigatus* (AY048754) and *Penicillium digitatum* (D78154) were used as outgroups of phylogenetic analysis. Interestingly, the result of the phylogenetic analysis of the partial β-tubulin gene showed that *M. pilosus* BCRC31503, *M. purpureus*, and *M. kaoliang* were placed into the same clade (**Figure 7**). This finding was in agreement with the grouping in Southern hybridization pattern by the presence or absence of MRT elements. Therefore, *M. pilosus* BCRC31503 may have been misidentified and should be reconsidered as *M. purpureus*. Since *M. pilosus*, *M. ruber*, and *M. barkeri* could not be differentiated using the partial β-tubulin genes, three species have been suggested to be synonymous, meaning that they should be classified as a single species. DNA hybridization among *Monascus* species (our unpublished data) also supported the identity of *M. pilosus* and *M. ruber*. The other genetic markers have been found to distinguish *Monascus* species using the D1/D2 region of the large subunit (LSU) rRNA genes (Park and Jong, 2003), and the ITS region (Park et al., 2004). These results also demonstrate that *M. ruber* and *M. pilosus* could not be differentiated. Moreover, the MRT non-LTR retrotransposons were widely distributed over *M. pilosus* and *M. ruber*. Hence, the two species were determined to be synonymous. In the taxonomy of *Monascus* (Hawksworth and Pitt, 1983), *M. barkeri* is under the name of *M. ruber.* The results herein were consistent with their theory. Furthermore, *M. sanguineus* was placed on a branch that was separate from *M. pilosus*, *M. ruber*, and *M. barkeri,* while the difference of phylogenetic distance corresponded to the results of Southern

hybridizations with varying intensities of bands of *M. pilosus*, *M. ruber*, and *M. barkeri*.

According to the phylogenetic subgroups established with the partial β-tubulin gene, the species were grouped by the presence or absence of MRT elements in the hybridization pattern (**Figure 7**). Since the MRT element was only detected in *M. pilosus*, *M. ruber*, *M. barkeri* and *M. sanguineus*, the MRT element was suggested to be present in the ancestors of the *Monascus* species, and absent from most other species or had diverged from *M. pilosus*. This phenomenon may have been caused by the genetic drift that is itself associated with small effective populations, which are responsible for an increase in the numbers of copies of elements in some species and a decline in others (Cizeron et al., 1998; Le Rouzic and Capy, 2005). However, horizontal transfer (HT) has been observed with the variable distribution of retrotransposons between different classes, phyla or kingdoms. The horizontal transfer is believed to be present in members of RTE clade (Župunski et al., 2001). Bov-B LINEs of the RTE clade exhibit a much low divergence between Ruminantia and Squamata, strongly indicating horizontal transfer. In this study, the result with the various species of *Monascus* revealed that MRT non-LTR retrotransposon was restricted to *M. sanguineus*, *M. pilosus*, *M. ruber*, and *M. barkeri*, of which the last three species were synonymous. These observations suggested that the MRT element was introduced into the ancestor of the *Monascus* species, and then diverged into *M. sanguineus*, *M. pilosus*, *M. ruber* and *M. barkeri*. Although the evidence for horizontal transfer events is presnet in non-LTR retrotransposons of RTE clade, different taxonomic groups must be sampled to determine whether the horizontal transfer events also occurred in the Tad1 clade.

Transposable elements may be regarded as genetic components that evolve in an ecological community of the host genome (Brookfield, 1995). During evolution, the

spread and distribution of the elements depend on not only their ability to amplify, but also the complex interactions between various families of elements, and between the elements and the hosts (Tu et al., 1998). Analyses of endogenous retrotransposable elements in *Monascus* species can be potentially used as markers in genetic mapping and in population studies. MRT was consistent with CgT1, which can distinguish between two groups based on presence or absence of non-LTR retrotransposon (He et al., 1996). Practically, our findings indicate that MRT has great potential in strain characterization and in studies of population structure and evolution in *Monascus* spp.

References

(1) Baba, S.; Kajikawa, M.; Okada, N.; Kawai., G. Solution structure of an RNA stem-loop derived from the 3' conserved region of eel LINE UnaL2. *RNA*. **2004**. *10*, 1380-1387.

(2) Bingle, L.E.H.; Simpson, T.J.; Lazarus, C.M. Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet. Biol.* **1999**. *26*, 209-223.

(3) Blesa, D.; Gandía, M.; Martínez-Sebastián, M.J. Distribution of the *bilbo* non-LTR retrotransposon in Drosophilidae and its evolution in the *Drosophila obscura* species group. *Mol. Biol. Evol.* **2001**, *18*, 585-592.

(4) Brookfield, J.F. Molecular evolution. Retroposon revivals. *Curr. Biol.* **1995**, *5*, 255-256.

(5) Burke, W.D.; Malik, H.S.; Rich, S.M.; Eickbush, T.H. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* **2002**, *19*, 619-630.

(6) Cambareri, E.B.; Helber, J.; Kinsey, J.A. *Tad1-1*, an active LINE-like element of *Neurospora crassa*. *Mol. Gen. Genet.* **1994**, *242*, 658-665.

(7) Casaregola, S.; Neuvéglise, C.; Bon, E.; Gaillardin, C. Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. *Mol. Biol. Evol.* **2002**, *19*, 664-677.

(8) Cizeron, G.; Lemeunier, F.; Loevenbruck, C.; Brehm, A.; Biémont., C. Distribution of the retrotransposable element 412 in *Drosophila* species. *Mol. Biol. Evol.* **1998**, *15*, 1589-1599.

(9) Daboussi, M.J.; Capy, P. Transposable elements in filamentous fungi. *Annu. Rev. Microbiol.* **2003**, 57, 275-299.

(10) Eickbush, T.H. Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* **1992**, *4*, 430-440.

(11) Endo, A. Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* (Tokyo) **1979**, *32*, 852-854.

(12) Flavell, A.J. Retroelements, reverse transcriptase and evolution. *Comp. Biochem. Physiol.* **1995**, *110*, 3-15.

(13) Goodwin, T.J.; Ormandy, J.E.; Poulter, R.T. L1-like non-LTR retrotransposons in the yeast *Candida albicans*. *Curr. Genet.* **2001**, *39*, 83-91.

(14) Gordon, D.; Desmarais, C.; Green, P. Automated finishing with autofinish. *Genome Res.* **2001**, *4*, 614-625.

(15) Goyon, C., Rossignol, J.L.; Faugeron, G. Native DNA repeats and methylation in *Ascobolus*. *Nucleic. Acids. Res.* **1996**, 24, 3348-3356.

(16) Hamer, J.E.; Farrall, L.; Orbach, M.J.; Valent, B.; Chumley, F.G. Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. *Proc. Natl. Acad. Sci. USA*. **1989**, *86*, 9981-9985.

(17) Hawksworth, D.L.; Pitt, J.I. A new taxonomy for *Monascus* species based on cultural and microscopical characters. *Aust. J. Botany*. **1983**, *31*, 51-65.

(18) He, C.; Nourse, J.P.; Kelemu, S.; Irwin, J.A.; Manners, J.M. *Cg*T1: a non-LTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. *Mol. Gen. Genet.* **1996**, *252*, 320-331.

(19) Le Rouzic, A.; Capy, P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics.* **2005**, *169*, 1033-1043.

(20) Maita, N.; Anzai, T.; Aoyagi, H.; Mizuno, H.; Fujiwara, H. Crystal structure of the endonuclease domain encoded by the telomere-specific long interspersed nuclear element, TRAS1. *J. Biol. Chem.* **2004**, *279*, 41067-41076.

(21) Malik, H.S; Burke, W.D.; Eickbush, T.H. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **1999**, *16*, 793-805.

(22) Murata, H.; Miyazaki, Y.; Yamada, A. *marY2N*, a LINE-like non-long terminal repeat (non-LTR) retroelement from the ectomycorrhizal homobasidiomycete *Tricholoma matsutake*. *Biosci. Biotechnol. Biochem.* **2001**, *65*, 2301-2305.

(23) Park, H.G.; Stamenova, E.K.; Jong, S.-C. Phylogenetic relationships of *Monascus* species inferred from the ITS and the partial β-tubulin gene. *Bot. Bull. Acad. Sin.* **2004**, *45*, 325-330.

(24) Park, H.G.; Jong, S.-C. Molecular characterization of *Monascus* strains based on the D1/D2 regions of LSU rRNA genes. *Mycoscience.* **2003**, *44*, 25-32.

(25) Peterson, D.G.; Tomkins, J.P.; Frisch, D.A.; Wing, R.A.; Paterson, A.H. Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. J. Agric. Genomics. Volume 5. **2000**.

(26) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406-425.

(27) Sambrook, J.; Fritsch, E.; Maniatis, T. Molecular cloning: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, New York. **1989**.

(28) Song, Z.; Krishna, S.; Thanos, D.; Strominger, J.L.; Ono, S.J. A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved X-box motif of the human major histocompatibility complex class II genes via a repeated Cys-His domain and functions as a transcriptional repressor. *J. Exp. Med.* **1994**, *180*, 1763-1774.

(29) Tu, Z.; Isoe, J.; Guzova, J.A. Structural, genomic, and phylogenetic analysis of Lian, a novel family of non-LTR retrotransposons in the yellow fever mosquito. *Aedes aegypti*. *Mol. Biol. Evol.* **1998**, *15*, 837-853.

(30) Wu, T.S.; Skory, C.D., Horng, J.S.; Linz, J.E. Cloning and functional analysis of

a beta-tubulin gene from a benomyl resistant mutant of *Aspergillus parasiticus*. *Gene.* 1996, 182, 7-12.

(31) Xiong, Y.; Eickbush, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **1990**, *9*, 3353–3362.

(32) Župunski, V.; Gubenšek, F.; Kordiš, D. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.* **2001**, *18*, 1849-1863.

Table 1. Strains used and GenBank accession numbers for the β-tubulin gene

Strain	Species[a]	MRT non-LTR retrotransposon[b]	Accession number of β-tubulin gene
BCRC 38072 (Taiwan isolate)	*Monascus pilosus*	+	DQ299886
BCRC 31502 (ATCC 16363)	*Monascus pilosus*, Type	+	AY498596
BCRC 31503 (ATCC 16364)	*Monascus pilosus*	−	DQ299887
BCRC 31533 (ATCC 16246)	*Monascus ruber*, Type	+	AY498589
BCRC 31523 (ATCC 16378)	*Monascus ruber*	+	DQ299888
BCRC 31534 (ATCC 16366)	*Monascus ruber*	+	AY498587
BCRC 31535 (ATCC 18199)	*Monascus ruber*	+	DQ299889
BCRC 33314 (ATCC 16371)	*Monascus ruber*	+	AY498588
BCRC 33323 (ATCC 18199)	*Monascus ruber*	+	DQ299890
BCRC 31542 (ATCC 16365)	*Monascus purpureus*, Type	−	DQ299891
BCRC 31541 (ATCC 16379)	*Monascus purpureus*	−	AY498598
BCRC 31615 (DSM 1379)	*Monascus purpureus*	−	DQ299892
BCRC 33325 (IFO 30873)	*Monascus purpureus*	−	DQ299893
BCRC 31506 (CBS 302.78)	*Monascus kaoliang*, Type	−	DQ299894
BCRC 33446 (ATCC 200613)	*Monascus sanguineus*, Type	+	AY498602

BCRC 33309 (ATCC 16966)	*Monascus barkeri*	+	DQ299895
BCRC 33310 (IMI 282587)	*Monascus floridanus*, Type	−	DQ299896
BCRC 33640 (ATCC 204397)	*Monascus lunisporas*, Type	−	AY498604
BCRC 33641 (ATCC 200612)	*Monascus pallens*, Type	−	AY498601

[a] "Type" indicates type strain.

[b] +, presence of MRT non-LTR retrotransposon; −, absence of MRT non-LTR retrotransposon.

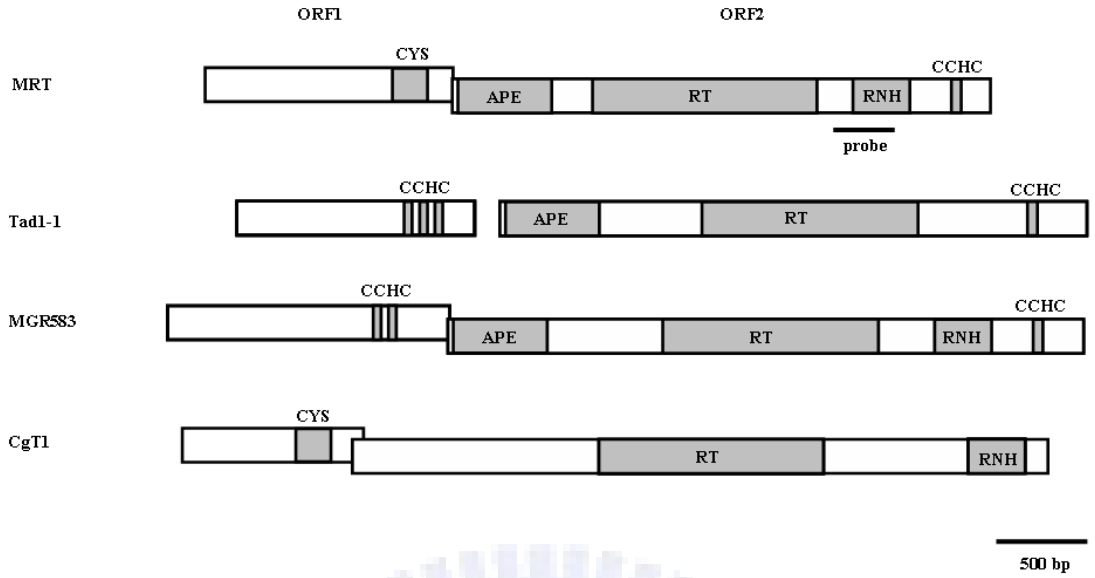


Figure 1. Structure of the MRT element. Schematic representation of the structural organization of the filamentous fungi non-LTR retrotransposons. Tad1-1, MGR583 and CgT1 in *Neurospora crassa*, *Magnaporthe grisea* and *Colletotrichum gloeosporioides* were obtained from the GenBank database using the following accession numbers, L25662, AF018033 and L76205, respectively. Southern hybridization analysis of MRT in the genomes of *Monascus* species hybridized with the probe indicated by small black bar. The abbreviation of CYS indicated the cysteine-rich region, RNH indicated the RNaseH, APE indicated the apurinic/apyrimidinic endonuclease domain, RT indicated the reverse transcriptase domain, and CCHC indicated the Cys-His region.

Figure 2. Deduced Amino acid sequences alignment of the MRT elements ORF1. The cysteine-rich nucleotide-binding domain represented a zinc finger with the consensus sequence $CX_6HXCX_6CHX_2HX_6C$ shown boxed.

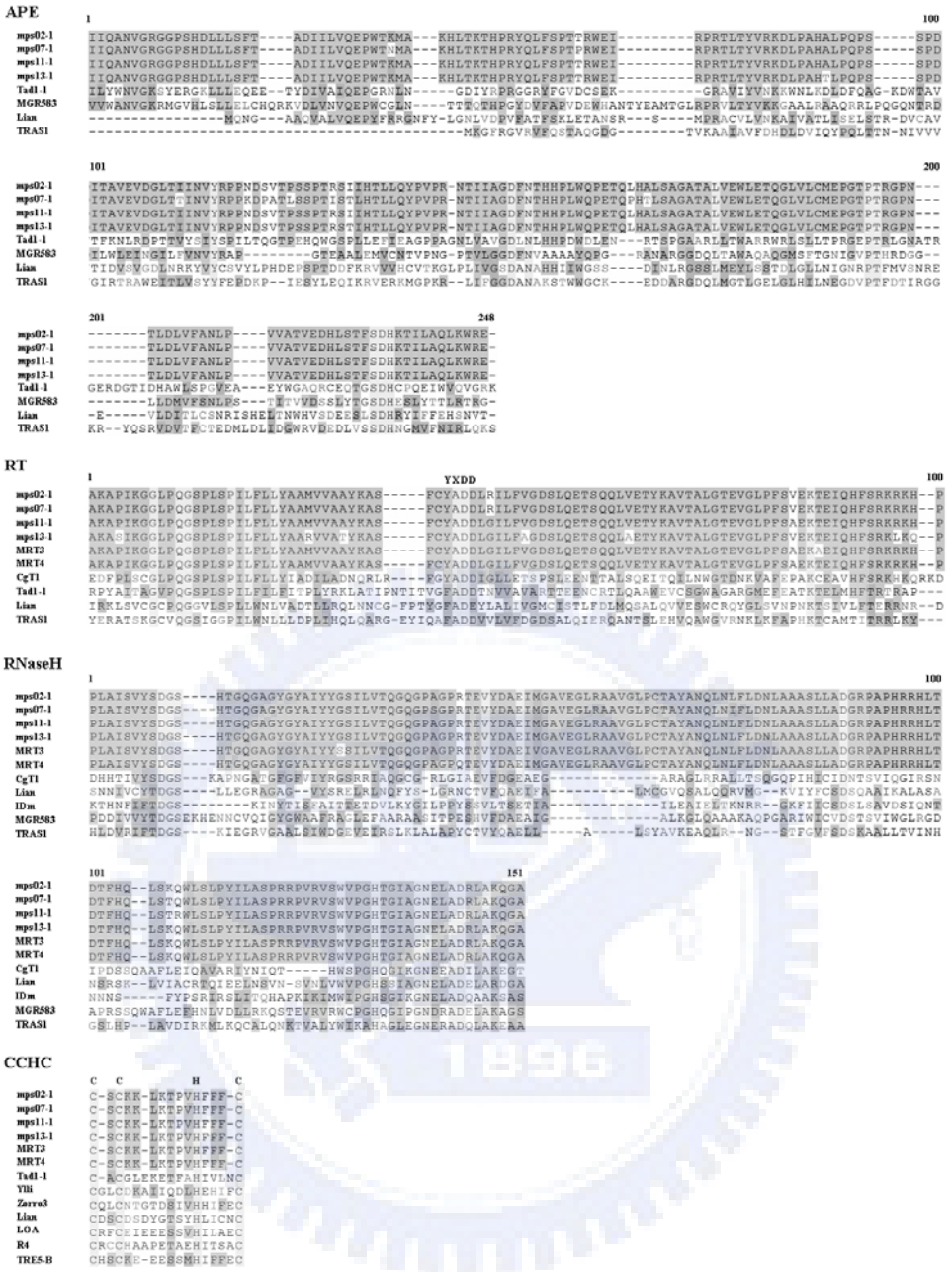


Figure 3. Multiple alignment of deduced amino acid sequences of the MRT elements with related organisms. Comparison of the N-termial apurinic/apyrimidinic endonuclease (APE) domains based on an alignment of approximately 210 amino acid residues. Comparison of the reverse transcriptase (RT) domain. Conserved YXDD residues, active site of reverse transcriptase domain, were indicated above the alignment. Comparison of the RNaseH domains based on an alignment of approximately 130 amino acid residues. Conserved residues of RNaseH were

described by Malik et al. (1999). Comparison of the C-terminal Cys-His region (CCHC). Conserved $CX_1CX_7HX_3C$ residues, putative zinc finger of the MRT element, were indicated above the alignment.

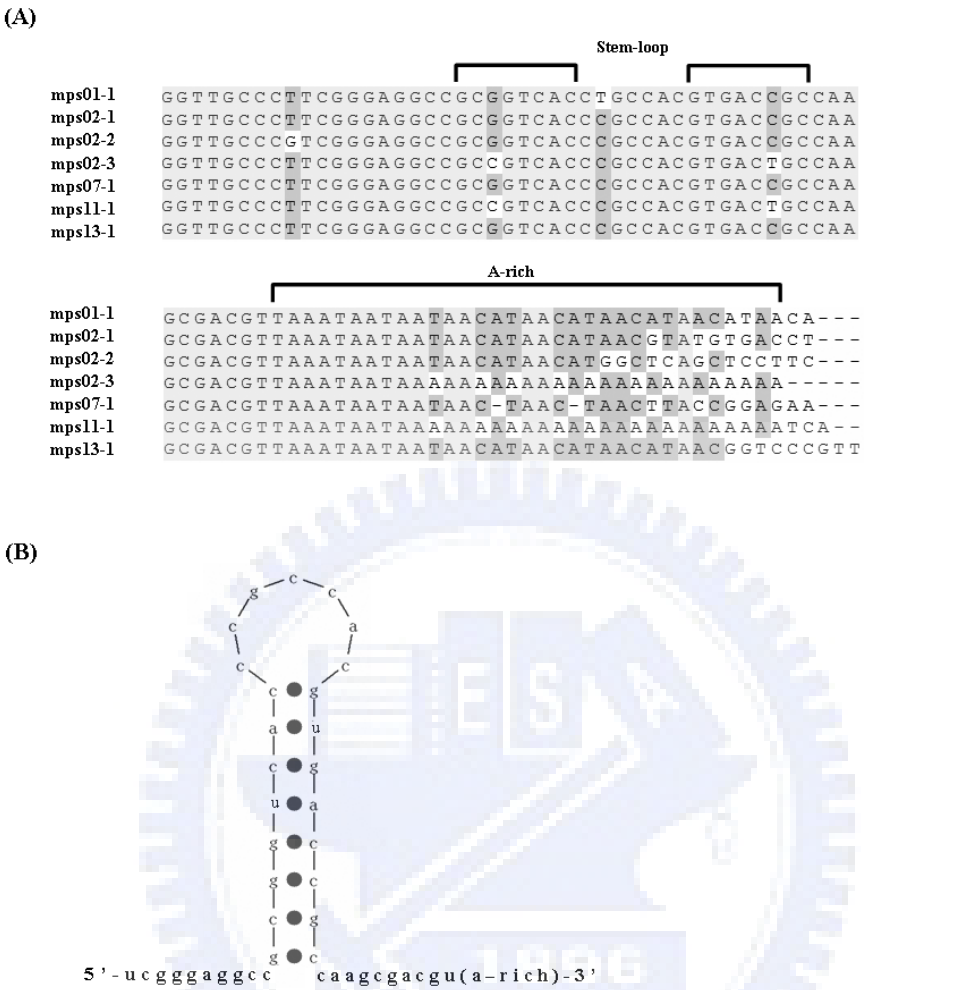


Figure 4. (A) Comparison of 3' conserved region of the MRT elements from BACs. The A-rich sequences and stem-loop were marked. (B) Putative secondary structure of 3' conserved region of the MRT element RNA as indicated in (A) was depicted.

(C) Sequences at the ends of the MRT elements. Four elements from BACs were shown. Putative target site duplications were indicated by underlining.

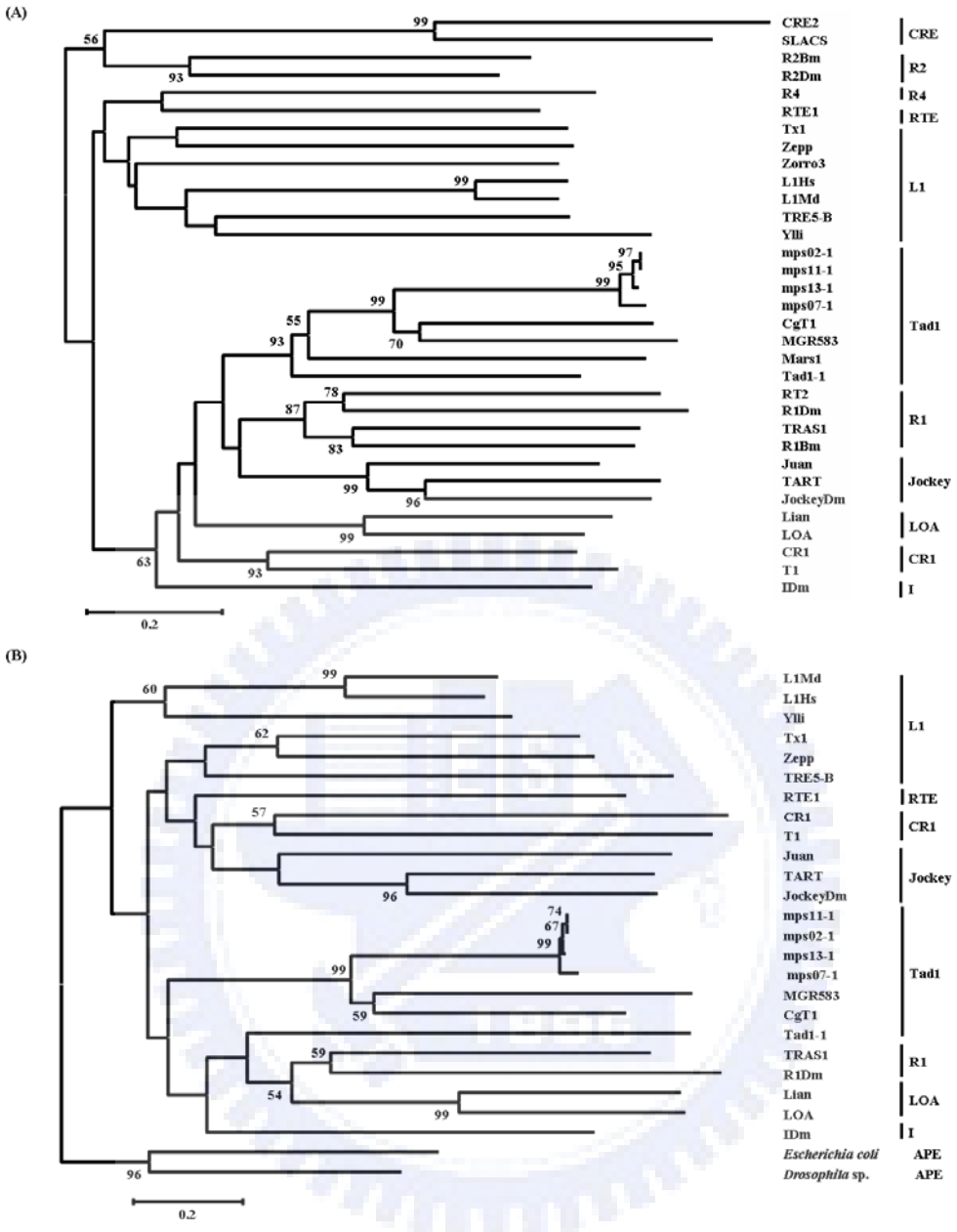


Figure 5. Phylogenetic tree of non-LTR retrotransposons from *M. pilosus* and various organisms. (A) The phylogeny of non-LTR retrotransposons based on the 11 conserved blocks of the reverse transcriptase domains defined by Malik et al. (1999) was constructed. The nucleotide sequences of MRT non-LTR retrotransposons were used in this study under the accession numbers DQ299897 to DQ299900. The tree was rooted using RT sequences of bacterial and fungal group II introns with *Lactococcus lactis* (P0A3U0), *Escherichia coli* (NP053121), *Sinorhizobium meliloti*

(CAC49872) *Neurospora crassa* (NP041729), *Saccharomyces cerevisiae* (NP009310), and *Schizosaccharomyces pombe* (S78199). (B) The phylogeny of non-LTR retrotransposons based on an alignment of approximately 210 amino acid residues of apurinic/apyrimidinic endonuclease (APE) domain. Accession numbers for the apurinic/apyrimidinic endonuclease used as the outgroup following: *Drosophila* sp. (AAB19427), and *Escherichia coli* (P09030). Bootstrap values were shown in the nodes according to the 1000 replications. Only bootstrap values >50 were shown. The tree was constructed by the neighbor-joining method (Saitou and Nei, 1987).

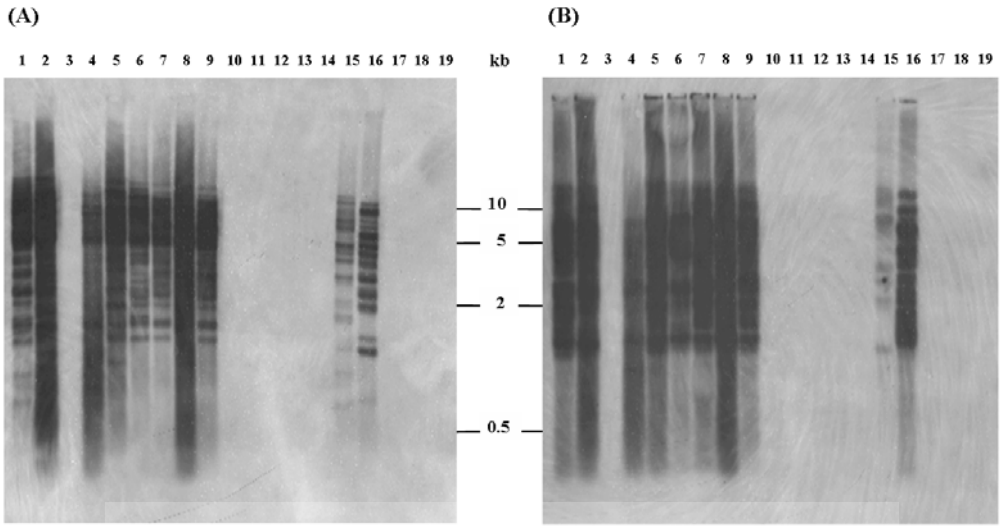


Figure 6. Southern hybridizations analyses of the MRT elements. Chromosome DNAs extracted from *Monascus* species were digested by *Eco*RI (A) and *Bam*HI (B) separated on electrophoresis gel and hybridized respectively with 422-bp probe. *Monascus* species— lane 1: *M. pilosus* BCRC38072; lane 2: *M. pilosus* BCRC31502; lane 3: *M. pilosus* BCRC31503; lane 4: *M. ruber* BCRC31533; lane 5: *M. ruber* BCRC31523; lane 6: *M. ruber* BCRC31534; lane7: *M. ruber* BCRC31535; lane 8: *M. ruber* BCRC33314; lane 9: *M. ruber* BCRC33323; lane 10: *M. purpureus* BCRC31542; lane 11: *M. purpureus* BCRC31541; lane 12: *M. purpureus* BCRC31615; lane 13: *M. purpureus* BCRC33325; lane 14: *M. kaoliang* BCRC31506; lane 15: *M. sanguineus* BCRC33446; lane 16: *M. barkeri* BCRC33309; lane 17: *M. floridanus* BCRC33310; lane 18: *M. lunisporas* BCRC33640; lane 19: *M. pallens* BCRC33641.

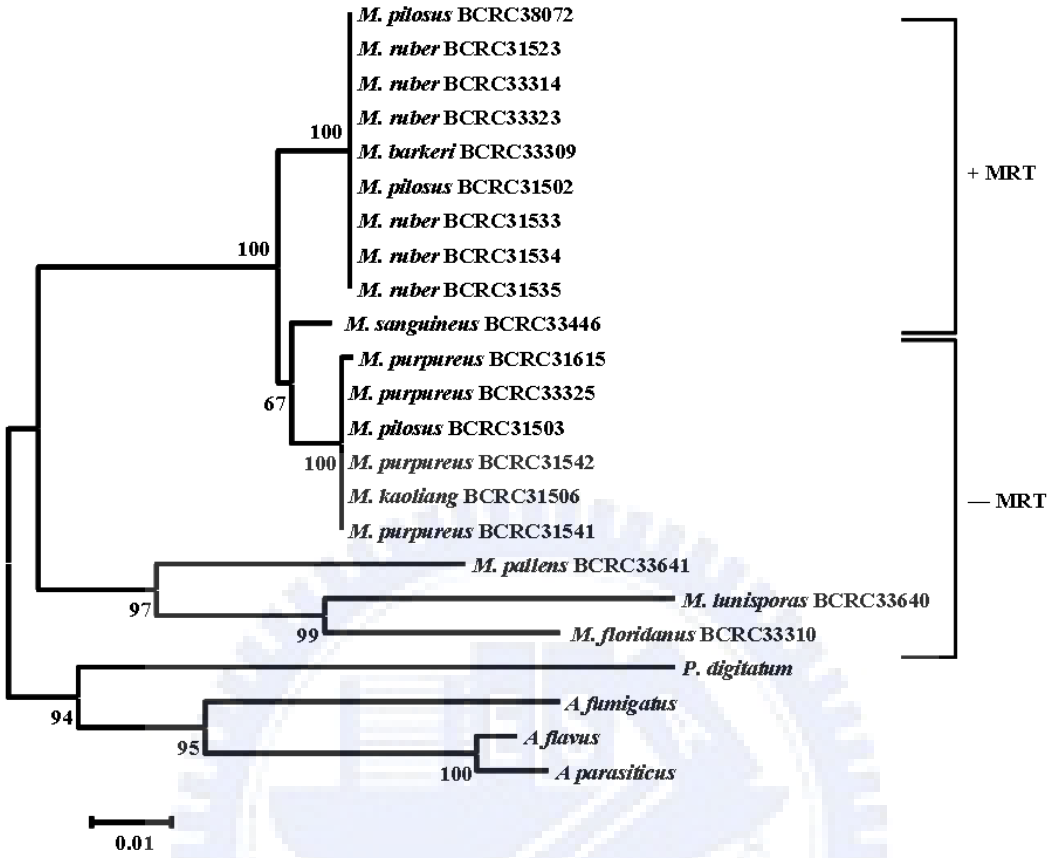


Figure 7. Phylogeny of *Monascus* species based on the partial β-tubulin gene amplified by PCR. The partial β-tubulin genes were used by the following accession numbers, DQ299886 to DQ299896, AY498587 to AY498589, AY498596, AY498598, AY498601, AY498602 and AY498604. Accession numbers for the β-tubulin genes were used as the outgroup following: *Aspergillus flavus* (M38265), *Aspergillus parasiticus* (L49386), *Aspergillus fumigatus* (AY048754), and *Penicillium digitatum* (D78154). Bootstrap values were shown in the nodes according to the 1000 replications. Only bootstrap values >50 were shown. The tree was constructed by the neighbor-joining method (Saitou and Nei, 1987). ＋ MRT: presence of MRT non-LTR retrotransposon. － MRT: absence of MRT non-LTR retrotransposon.

Published papers in this study

1. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Wang, C.-L., Yuan, G.-F. Characterization of MRT, a new non-LTR retrotransposon in *Monascus* spp. Bot. Bull. Acad. Sinica. 2007. 48, 377-385.
2. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Wang, C.-L., Chen, I.-C., Wu, W.-J., Wu, M.-D., Yuan, G.-F. Cloning and characterization of monacolin K biosynthetic gene cluster from *Monascus pilosus*. J. Agric. Food Chem. 2008. 56, 5639-5646.
3. **Chen, Y.-P.**, Chen, I.-C., Hwang, I.-E., Yuan, G.-F., Liaw, L.-L., Tseng, C.-P. Selection of an effective red-pigment producing *Monascus pilosus* by efficient transformation with aurintricarboxylic acid. 2008. Biosci. Biotechnol. Biochem. (Acepted)
4. **Chen Y.-P.**, Tseng C.-P., Chien I.-L., Wang W.-Y., Liaw L.-L., Yuan G.-F. Exploring the distribution of citrinin biosynthesis related genes among *Monascus* species. 2008. J. Agric. Food Chem. (Acepted)
5. **Chen Y.-P.**, Tseng C.-P., Wang W.-Y., Lin, Y.-S., Liaw L.-L., Yuan G.-F. Overexpression of *mokH* gene encoding transcription factor of monacolin K biosynthesis gene cluster in *Monascus pilosus*. 2008. (Manuscript in preparation)

Patent

	Title of invention	Country	Patent Number	Inventor	patentee
1	膽固醇合成抑制劑 Monacolin K 生合成相關基因	ROC	I297358	**陳煜沛等**	食品研究所
2	反轉錄轉座子以及其應用	ROC	I296010	**陳煜沛等**	食品研究所
3	Monacolin K Biosynthesis Genes	Korea	10-0662700	**陳煜沛等**	食品研究所

Conference paper

1. **Chen, Y.-P.**, Liaw, L.-L., Wu, M.-D., Wang, C.-L, Tseng, C-P., Yuan, G.-F. 2005. Cloning, characterization and expression of a polyketide synthase gene involved in monacolin k biosynthesis from *Monascus* sp. (23rd Fungal Genetics Conference. U.S.A) (International Conference)
2. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Wang, C.-L., Yuan, G.-F. 2005. Identification and phylogeny of the active MRT non-LTR retrotransposon in the filamentous fungus, *Monascus pilosus*. (中華民國真菌學會研討會.)
3. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Cheng, M.-J., Wang, C.-L., Yuan, G.-F. 2005. The biosynthetic gene cluster for the cholesterol-lowering agent monacolin

K in *Monacus pilosus*. (生物催化暨生物技術國際研討會.) (International Conference)

4. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Cheng, M.-J., Yuan, G.-F. 2006. Coexpression and characterization of polyketide synthase (*mkA*) and phosphopantetheinyl transferase (*sfp*) in *Escherichia coli*. (第十一屆生化工程研討會.)
5. **Chen, Y.-P.**, Tseng, C.-P, Liaw, L.-L., Lin, Y.-S., Wang, C.-L., Chen, I.-C., Wu, M.-D., Wu, W.-J., Yuan, G.-F. 2007. Cloning and Characterization of Monacolin K Biosynthetic Gene Cluster from *Monascus pilosus*. (NCTU and NTHU Biotech Conference, Hsin-Chu, Taiwan)
6. **Chen, Y.-P.**, Tseng, C.-P., Liaw, L.-L., Wang, C.-L., Chen, I.-C., Wu, W.-J., Wu, M.-D., Yuan, G.-F. 2008. Characterization of monacolin K biosynthetic gene cluster in *Monascus pilosus* BCRC38072 (中華民國真菌學會研討會.)