Computational analysis reveals industrial natural products and toxins in charcoal rot pathogen, *Macrophomina phaseolina*

Sharuti Thakur and Tarakanta Jana*

CSIR-National Institute of Science Communication and Information Resources, 14 Satsang Vihar Marg, New Delhi 110 067, India

The fungal polyketide synthases (PKS) are responsible for the biosynthesis of several polyketide natural products, mycotoxins, pigments, etc. In the present times, we use computational tools to gain insight into polyketide natural products that may contribute to the metabolic versatility of this important phytopathogenic filamentous fungi. In total, we have identified 17 type-I PKS related gene clusters from the *Macrophomina phaseolina* genome. Among these 27 ketosynthase (KS) domains have been retrieved and used for the study. The study reveals that genome of *M. phaseolina* comprises non-reducing (NR), partially reducing (PR) and reducing (R) type of polyketides, and are clustered into three clades and several subclades. The phylogenetic analysis of KS domain sequences of *M. phaseolina* indicates that some *PKS* sequences are most closely related to polyketide natural product homologs such as lovastatin diketide, mycotoxins (fumonisin, citrinin and patulin) and pigment melanin. We also found eight orphan KS domains from three reducing PKS, i.e. MPH10374, MPH10375 and MPH10376. The study represents a potential novel source of industrially important polyketide natural products.

Keywords: Computational tools, *Macrophomina phaseolina*, polyketide synthase, phylogenetics.

FUNGI are prolific producers of low-molecular-weight molecules with various biological activities, such as antibiotic, immunosuppressive, cholesterol-lowering and antitumour activities[1–5]. In addition, fungi produce natural dyes and mycotoxins which are of importance in dye and food industry[6,7]. However, bioinformatics analysis of the sequenced fungal genomes indicates that their potential to produce secondary metabolites is greatly underestimated, and it appears that many cryptic natural products await discovery[8–10]. Fungi have the potential to produce a far greater number of polyketides than the known polyketides isolated from them till date[11]. The biosynthesis of fungal polyketides is primarily governed by iterative type-1 polyketide synthases (*PKS*) genes. In the past decade, more knowledge about biosynthesis mechanisms of fungal PKS has been acquired. PKS generate molecular diversity by utilizing different starter molecules and by controlling the final length of the polyketide[12].

The charcoal rot pathogen *Macrophomina phaseolina* (Tassi) Goidanich, is a soil- and seed-borne plant pathogen, extremely well adapted to different ecological environments. It causes charcoal rot in more than 500 host plants[13–15]. The genome size of *M. phaseolina* is 49.25 Mb, showing 14,249 genes profusely rich in GC content[16]. In agriculture, filamentous plant pathogenic charcoal rot fungi are responsible for critical loss of crop plants. For example, in 2002 the United States suffered a total loss of US$ 178.80 million in soybean alone[17]. In Bangladesh, the fibre yield of jute is reduced by 30% due to this pathogen. This pathogen causes huge economic losses in India and other places as well[18]. Due to the perception of this fungi as a plant pathogen, we tend to neglect its beneficial role as producers of industrially valuable products. The phytopathogen has remarkably diverse metabolism capable of catabolizing a wide variety of plant metabolites, withstand abiotic stresses (soil, high temperature and drought) and biotic stresses (resource-competitive soil rhizobacteria, antagonistic fungi and other organisms in soil), presumably producing a wide variety of useful polyketides for mankind, of which no detailed study has been done so far. Many computational tools and methodologies are used for the identification, classification, dereplication and prioritization of PKS in the genomes[19–29]. The present study is aimed at identifying type-1 PKS of *M. phaseolina* genome through genome mining computational tools in order to explore the differences in PKS repertoire. We have taken a phylogenomic approach to study ketosynthase (KS) encoded in the genome of *M. phaseolina* with other filamentous fungi, bacteria and other organisms.

Materials and methods

Genome sequence

Whole genome sequence of *M. phaseolina* MS6 was retrieved from GenBank with accession no. AHHD 00000000.1dd (ref. 16). Contigs of draft genome were extracted and saved as a fasta file.

*For correspondence. (e-mail: tkj@niscair.res.in)

A web-based tool antiSMASH (antibiotics and Secondary Metabolite Analysis Shell), version 3.0 which is a secondary metabolite gene cluster prediction and analysis tool was used in the present study[30,31]. It is the most comprehensive tool for identification and analysis of secondary metabolite biosynthetic gene clusters in fungal and bacterial genome data. AntiSMASH also implements other tools for domain identification of PKS.

GenBank accession number was given as input for antiSMASH or extracted fasta file of whole-genome sequence of *M. phaseolina* can be uploaded to antiSMASH. The predicted secondary metabolite gene clusters from antiSMASH consist of PKS, nonribosomal peptide synthetase (NRPS), hybrid PKS/NRPS, terpenes, hglKS and hglKS/PKS. The clusters responsible for biosynthesis of PKS in *M. phaseolina* have been considered for this study. AntiSMASH output contains domain composition and information about the domains of PKS. The KS domains from antiSMASH output were further explored and amino acid sequences of all the KS domains of *M. phaseolina* were retrieved for further study. In addition, 120 PKS protein sequences characterized by Kroken *et al.*[32] were obtained from GenBank (see Supplementary Information online, Note S1). Among these, 23 PKS sequences were taken from *Cochliobolus heterostrophus*, 19 PKS from *Botryotinia fukeliana*, 14 PKS from *Gibberella moniliformis*, 16 PKS from *Gibberella zea* and 48 sequences from other fungi and bacteria (Table 1).

Characterization of KS in PKS

All the domain amino acid sequences of *M. phaseolina* from antiSMASH output were retrieved. The amino acid sequences of KS domain of 17 *PKS* gene clusters were used to generate multiple alignments and phylogenetic tree. Multiple sequence alignment was conducted using CLUSTAL X2 program (ref. 33). Phylogenetic trees were constructed using distance method from PHYLIP package version 3.695 (refs 34, 35). For the distance method, a distance matrix was constructed using the PROTDIST program and then the matrix was transformed into a tree by NEIGHBOR program. In order to verify the accuracy of the tree, multiple datasets were generated using the SEQBOOT program with 1000 bootstrap replicates. A tree was built from each replicate with PROTDIST then bootstrap values were computed with CONSENSE program[36].

Prediction of polyketide natural products PKS in *M. phaseolina*

GenBank accession numbers of 120 PKS obtained from fungi, bacteria and eukaryotes[32], and 17 PKS of *M. phaseolina* were given as input for antiSMASH or extracted fasta file were uploaded to antiSMASH and KS domain sequences from all the PKS were retrieved. The phylogenetic tree was constructed with KS sequences of *M. phaseolina* PKS to examine if there is any similarity of known polyketide natural products. Neighbour Joining (NJ) method was used, i.e. NEIGHBOR in PHYLIP with 1000 bootstrapping steps and the resulting consensus tree constructed. Bootstrapping was performed to assess the robustness of the phylogeny. The phylogenetic trees were visualized and edited Treeview.

Results and discussion

Analyses of type-I PKS

In the present study, the classification of PKS was based on information in the literature and domain composition predicted by computational tools. In total, we identified 17 type-I PKS-related gene clusters for the *M. phaseolina* genome. Therefore, we assigned 6-MSAS (6-methylsalicylic acid synthase) type to PKS with domain order KS–AT–DH–KR–ACP. The remaining type-1 PKS were assigned to non-reducing (NR), partially reducing (PR) and reducing (R) type depending upon the existence of reducing domain and information in literature (Table 2). In the present study, NR-PKS contain a core set of domains AT, KS, ACP and TE/CYC. In addition, malonyl acyl transferase (MAT) and product template (PT) domains are also present, NR-PKS do not contain any reducing domains like DH, KR or ER. PR-PKS contain KR and possibly DH domain in addition to the core set of domains. PR-PKS do not have ER, starter unit : acyl-carrier protein transacylase (SAT), PT or TE domain. Reducing PKS contain KR, DH and ER reductive domains in addition to the core set of domains. All the detected PKS encode KS protein, but one PKS at locus MPH13401 is hypothetical protein. The KS domains are the most conservative domains of PKS. Figure 1 shows the results of phylogenetic analysis.

Analyses of KS domain

Phylogenetic analysis of KS revealed that there are three clades (clades 1–3; Figure 1). Three PKS at locus MPH09202, MPH02766 and MPH09266 did not fall into any clade because they did not cluster with any other PKS. Among these three PKS, two at locus MPH09202 and MPH02766 were non-reducing and had exact domain order, i.e. KS–AT–ACP–cMT. The remaining PKS at MPH09266 had domain organization as KS–AT–DH–ACP–ACP–cMT; because of the gain of one reducing domain DH, it became partially reducing.

Clade 1 consists of three PKS which were present at loci MPH10374, MPH10375 and MPH10376; these were found to be reducing, partially reducing and

Table 1. GenBank accession numbers for previously published sequences included in the present study

Species	GenBank accession no.	Gene
Cochliobolus heterostrophus	AY495643.1	Polyketide synthase (*PKS2*)
C. heterostrophus	AY495644	Polyketide synthase (*PKS3*)
C. heterostrophus	AY495645	Polyketide synthase (*PKS4*)
C. heterostrophus	AY495646	Polyketide synthase (*PKS5*)
C. heterostrophus	AY495647	Polyketide synthase (*PKS6*)
C. heterostrophus	AY495648	Polyketide synthase (*PKS7*)
C. heterostrophus	AY495649	Polyketide synthase (*PKS8*)
C. heterostrophus	AY495650	Polyketide synthase (*PKS9*)
C. heterostrophus	AY495651	Polyketide synthase (*PKS10*)
C. heterostrophus	AY495652	Polyketide synthase (*PKS11*)
C. heterostrophus	AY495654	Polyketide synthase (PKS12)
C. heterostrophus	AY495655	Polyketide synthase (PKS14)
C. heterostrophus	AY495656	Polyketide synthase (*PKS15*)
C. heterostrophus	AY495657	polyketide synthase (*PKS16*)
C. heterostrophus	AY495658	Polyketide synthase (*PKS17*)
C. heterostrophus	AY495659	Polyketide synthase (*PKS18*)
C. heterostrophus	AY495660	Polyketide synthase (*PKS19*)
C. heterostrophus	AY495661	Polyketide synthase (*PKS20*)
C. heterostrophus	AY495662	Polyketide synthase (*PKS21*)
C. heterostrophus	AY495663	Polyketide synthase (*PKS22*)
C. heterostrophus	AY495664	Polyketide synthase (*PKS23*)
C. heterostrophus	AY495665	Polyketide synthase (*PKS24*)
C. heterostrophus	AY495666	Polyketide synthase (*PKS25*)
Botryotinia fuckeliana	AY495606	Polyketide synthase (*PKS1*)
B. fuckeliana	AY495607	Polyketide synthase (*PKS2*)
B. fuckeliana	AY495608	Polyketide synthase (*PKS3*)
B. fuckeliana	AY495609	Polyketide synthase (*PKS4*)
B. fuckeliana	AY495610	Polyketide synthase (*PKS5*)
B. fuckeliana	AY495611	Polyketide synthase (*PKS6*)
B. fuckeliana	AY495613	Polyketide synthase (*PKS8*)
B. fuckeliana	AY495614	Polyketide synthase (*PKS9*)
B. fuckeliana	AY495615	Polyketide synthase (*PKS10*)
B. fuckeliana	AY495616	Polyketide synthase (*PKS11*)
B. fuckeliana	AY495617	Polyketide synthase (PKS12)
B. fuckeliana	AY495618	Polyketide synthase (*PKS13*)
B. fuckeliana	AY495619	Polyketide synthase (*PKS14*)
B. fuckeliana	AY495620	Polyketide synthase (*PKS15*)
B. fuckeliana	AY495621	Polyketide synthase (*PKS16*)
B. fuckeliana	AY495622	Polyketide synthase (*PKS17*)
B. fuckeliana	AY495623	Polyketide synthase (*PKS18*)
B. fuckeliana	AY495624	Polyketide synthase (*PKS19*)
B. fuckeliana	AY495625	Polyketide synthase (*PKS20*)
Gibberella moniliformis	AY495591	Polyketide synthase (*PKS1*)
G. moniliformis	AY495593	Polyketide synthase (*PKS3*)
G. moniliformis	AY495594	Polyketide synthase (*PKS4*)
G. moniliformis	AY495595	Polyketide synthase (*PKS5*)
G. moniliformis	AY495596	Polyketide synthase (*PKS6*)
G. moniliformis	AY495597	Polyketide synthase (*PKS7*)
G. moniliformis	AY495598	Polyketide synthase (*PKS8*)
G. moniliformis	AY495599	Polyketide synthase (*PKS9*)
G. moniliformis	AY495600	Polyketide synthase (*PKS10*)
G. moniliformis	AY495601	Polyketide synthase (*PKS11*)
G. moniliformis	AY495602	Polyketide synthase (*PKS12*)
G. moniliformis	AY495603	Polyketide synthase (*PKS13*)
G. moniliformis	AY495604	Polyketide synthase (*PKS14*)
G. moniliformis	AY495605	Polyketide synthase (*PKS15*)
Gibberella zeae	AY495626	Polyketide synthase (*PKS1*)
G. zeae	AY495627	Polyketide synthase (*PKS2*)
G. zeae	AY495628	Polyketide synthase (*PKS3*)
G. zeae	AY495629	Polyketide synthase (*PKS4*)
G. zeae	AY495630	Polyketide synthase (*PKS5*)
G. zeae	AY495631	Polyketide synthase (*PKS6*)

(*Contd*)

Table 1. (*Contd*)

Species	GenBank accession no.	Gene
G. zeae	AY495632	Polyketide synthase (*PKS7*)
G. zeae	AY495633	Polyketide synthase (*PKS8*)
G. zeae	AY495634	Polyketide synthase (*PKS9*)
G. zeae	AY495635	Polyketide synthase (*PKS10*)
G. zeae	AY495636	Polyketide synthase (*PKS11*)
G. zeae	AY495637	Polyketide synthase (*PKS12*)
G. zeae	AY495638	Polyketide synthase (*PKS13*)
G. zeae	AY495639	Polyketide synthase (*PKS14*)
G. zeae	AY495640	Polyketide synthase (*PKS15*)
G. zeae	AY495641	Polyketide synthase (*PKS16*)
Aspergillus fumigatus	AAC39471	Polyketide synthase (*alb1*)
A. nidulans	AAA81586	Polyketide synthase (*pksST*)
A. nidulans	CAA46695	Polyketide synthase for naphthopyrone YWA1
Aspergillus parasiticus	AAC41675	Polyketide synthase (*pksL1*)
Aspergillus terreus	BAB88688	Polyketide synthase (*at1*)
A. terreus	BAB88689	Polyketide synthase (*at4*)
A. terreus	BAB88752	Polyketide synthase (*at5*)
A. terreus	AAD39830	Lovastatin nonaketide synthase (*lov B*)
A. terreus	AAC49814	Polyketide synthase (*pksM*)
A. terreus	AAD34559	Polyketide synthase (*lovF*)
Bacillus subtilis	AAF08795	MycA
Bombyx mori	AAB53258	FAS p270
Byssochlamys nivea	AAK48943	6-Methylsalicylic acid synthase
Caenorhabditis elegans	NP_492417.2	FASN-1
Colletotrichum lagenaria	BAA18956	Polyketide synthase (*pks1*)
Glarea lozoyensis	AAN59953	Polyketide synthase 1 (*pks1*)
Gibberella moniliformis	AAD43562	Fum1p
Homo sapiens	AAC50259	FAS
Microcystis aeruginosa	AAF00957	McyG
M. aeruginosa	AAF00958	McyE
Monascus purpureus	CAC94008	Polyketide synthase 1 (*pks1*)
Mycobacterium leprae	AAA17364	pksE
Myxococcus xanthus	CAB38084	Ta1
Nodulisporium sp.	AAD38786	Melanin polyketide synthase
Nostoc sp. *GSV224*	AAF15891	NosB
Penicillium griseofulvum	AAB49684	Pks2
Penicillium citrinum	BAC20564	mlcA
P. citrinum	BAC20566	mlcB
Penicillium patulum A	CAA39295	6-Methyl salicylic acid synthase
Pseudomonas syringae	AAD03047	Type-I polyketide synthase (*cfa6*)
P. syringae	AF098795	Cfa7
Sorangium cellulosum	AAF26919	polyketide synthase (*epoA*)
S. cellulosum	AAF2692	polyketide synthase (*epoc*)
S. cellulosum	AAF62885	EpoF
Stigmatella aurantiaca	AAK57186	MxaB2
S. aurantiaca	AAK57187	MxaC
S. aurantiaca	AAK57187	MxaC2
S. aurantiaca	AAK57187	MxaC3
S. aurantiaca	AAK57188	MxaD
S. aurantiaca	AAK57189	MxaE
S. aurantiaca	AAK57190	MxaF
S. natalensis	CAC20930	PimS0 protein
S. natalensis	CAC20921	PimS2 protein module 2
S. natalensis	CAC20921	PimS2 protein module 3
Streptomyces viridochromogenes	AAK83194	aviM
Xanthomonas albilineans	AAK15074	Albicidin multifunctional polyketide-peptide synthase (xabB module 2)
X. albilineans	AAK15074	Albicidin multifunctional polyketide-peptide synthase (xabB module 3)
Xylaria sp.	AAM93545	PKS12

Table 2. Domain characteristics of polyketide synthases from *Macrophomina phaseolina*

Gene cluster no.	Locus tag	Domain organization	Degree of reduction
4	MPH02766	KS–AT–ACP–cMT	Non-reducing
9	MPH03673	KS–AT–DH–KR–ACP	6-MSAS
13	MPH04233	KS–AT–ACP	Non-reducing
16	MPH06435	KS	Non-reducing
	MPH06436	KS–AT–DH–ER–KR–ACP	Reducing
20	MPH07079	KS–AT–DH–cMT–ER–KR	Reducing
24	MPH07550	KS–AT–DH–cMT	Partially reducing
33	MPH09202	KS–AT–ACP–cMT	Non-reducing
34	MPH09266	KS–AT–DH–ACP–ACP–cMT	Partially reducing
38	MPH09899	KS	Non-reducing
40	MPH10194	KS–AT–DH–KR–ACP	6-MSAS
41	MPH10374	ER–cMT–KR–ACP ACP–ACP–ACP–KS–DH–ER–cMT–KR–ACP–KS–AT–DH–cMT–KR–ACP–KS–KR–ACP–KS	Reducing
41	MPH10375	ACP–KS–KR–ACP–KS–AT	Partially reducing
41	MPH10376	cMT–ACP–KS–AT–DH–cMT–KR–KS–DH–ER–cMT–KR–ACP	Reducing
45	MPH11432	KS–AT	Non-reducing
45	MPH11434	KS–AT–DH–cMT–TD	Partially reducing
45	MPH11441	KS	Non-reducing
46	MPH11504	KS–AT–ACP	Non-reducing
49	MPH12258	KS–AT–ACP–ACP–TE	Non-reducing
58	MPH13262	KS–AT–DH	Partially reducing
60	MPH13401	KS–AT–DH–ACP–TE	Partially reducing
61	MPH13431	KS–AT–DH–cMT–ER–KR–ACP	Reducing

AT, Acyl transferase; ACP, Acyl carrier protein; KS, Ketosynthase; KR, Ketoreductases; DH, Dehydratases; ER, Enoyl reductases; cMT, C-terminal methyl transferases; TE, Thioesterase; 6-MSAS, 6-Methylsalicylic acid synthase and TD, C-terminal domain.

non-reducing respectively. The domain organization of these PKS was found to be unusual. Due to their unusual nature, all these PKS fell under the same clade and had more than one KS and more than one ACP domains. PKS at locus MPH10374 was found to be reducing, because it contains all three reductive domains. The domain organization was ER–cMT–KR–ACP–ACP–ACP–ACP–KS–DH–ER–cMT–KR–ACP–KS–AT–DH–cMT–KR–ACP–KS–KR–ACP–KS, in that particular order. This unusual PKS contains seven ACP, one AT, two cMT, four KR, two DH, two ER, and four KS domains together at one locus only. To differentiate these four KS domains from one another, different names were used in the phylogenetic tree construction, i.e. MPH10374a, MPH10374b, MPH10374c and MPH10374d. PKS at MPH10375 was partially reducing and contained only one reducing domain KR, one AT domains, two ACP and two KS domains, while constructing the phylogenetic tree KS domains were named as MPH10375a and MPH10375b.

Clade 2 contains all types of PKS, i.e. non-reducing, partially reducing, reducing and 6-MSAS. Clade 2 was subdivided into three subclades. Subclade 1 had two reducing PKS. The PKS at locus MPH13431 did not cluster with any other KS domain and had domain organization KS–AT–DH–cMT–ER–KR–ACP. The second reducing PKS at locus MPH06436 had domain organization KS–AT–DH–ER–KR–ACP and was clustered with PKS at locus MPH06435, which had only one domain, i.e. KS domain; this domain is said to be incomplete because it lacks other core domains. In subclade 1 there were two other non-reducing PKS at loci MPH09899 and MPH11432; their KS domains were clustered with one another. PKS at locus MPH09899 had only KS domain therefore, it is an incomplete domain. PKS at locus MPH11432 had KS–AT. Subclade 2 had one reducing, one non-reducing and three partially reducing PKS. Two partially reducing PKS at loci MPH13262 and MPH07550 were found to be clustered together. The PKS at MPH 13262 had domain organization KS–AT–DH, while PKS at MPH07550 had domain organization KS–AT–DH–cMT. Another partially reducing PKS at locus MPH11434 did not cluster with any other PKS and had KS–AT–DH–cMT–TD domain organization. Similarly, PKS at locus MPH07079 did not cluster with any other PKS, having domain organization KS–AT–DH–cMT–ER–KR. Subclade 2 also contained one non-reducing PKS at locus MPH11441 loci with only one KS domain (incomplete domain). Subclade 3 had two 6-MSAS-type PKS at loci MPH03673 and MPH10194. Their KS domains were found to be clustered with one another and particular order of domains was KS–AT–DH–KR–ACP.

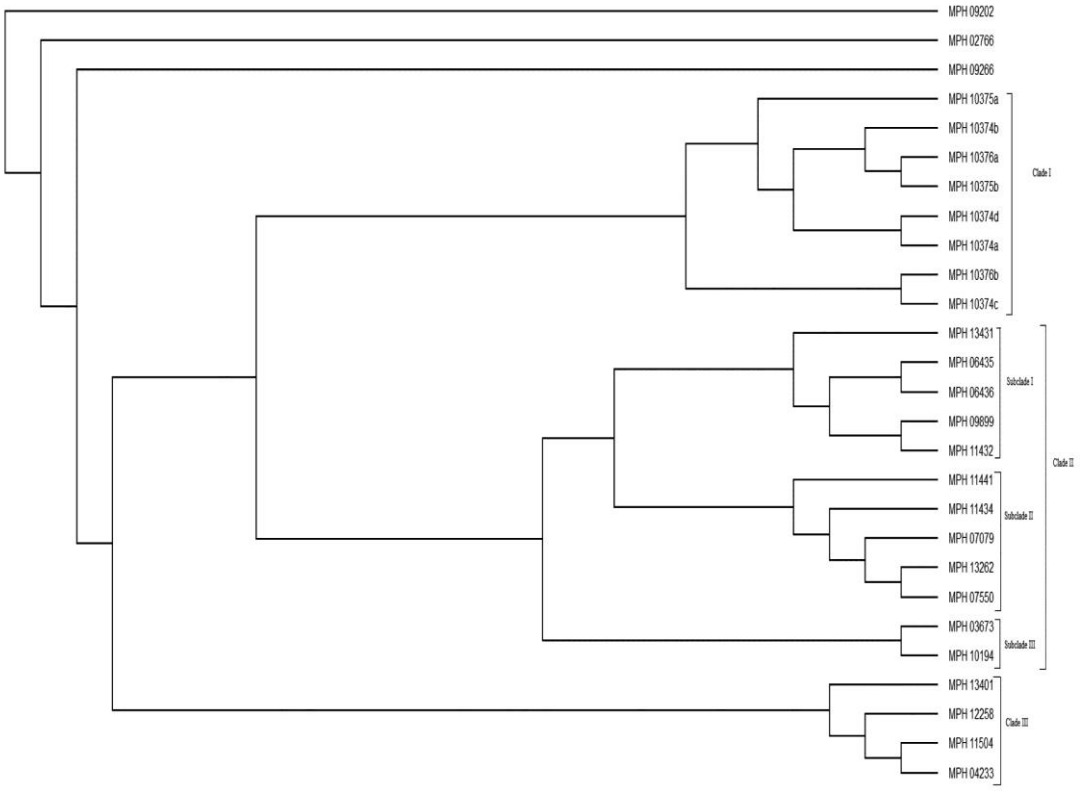

Figure 1. Neighbour joining (NJ) phylogentic tree of ketosynthase (KS) domains of type-1 polyketide synthases (PKS) using Phylip program, setting the bootstrap analysis to 1000 runs.

Clade 3 had three non-reducing and one partially reducing PKS. The partially reducing PKS at locus MPH13401 did not cluster with any other PKS and had domains in the order KS–AT–DH–ACP–TE. The non-reducing PKS at locus MPH12258 did not cluster with any other PKS and domains were organized as KS–AT–ACP–ACP–TE. The remaining two non-reducing PKS at loci MPH11504 and MPH04233 were clustered together, and both had domains organized in the particular order KS–AT–ACP.

Industrially important natural products and their cryptic gene clusters lying in M. phaseolina genome

The genome of *M. phaseolina* has 17 PKS encoding gene clusters containing 27 KS domains. The organizational framework for fungal PKS based on phylogeny of KS domain has been well established by Kroken *et al.*[32]. Therefore, we have analysed the position of *M. phaseolina* PKS proteins in the phylogenetic tree (Figure 2). In the present study we predicted a cholesterol-lowering agent (lovastatin), melanin pigment and three mycotoxins (citrinin, fumonisin and patulin) from the genome of *M. phaseolina*. These polyketide natural products have pharmaceutical, industrial and agricultural importance respectively.

Lovastatin: This is a polyketide metabolite produced by the filamentous fungus *Aspergillus terreus*. Lovastatin is a statin drug used for lowering cholesterol. It inhibits the rate-limiting step in the biosynthesis of cholesterol, the conversion of (3S)-hydroxymethylglutaryl-CoA (HMG-CoA) into mevalonate by the enzyme HMG-CoA reductase. Lovastatin therefore exhibits strong cholesterol lowering activity and is in clinical use for treatment of hypercholesterolemia. Lovastatin is composed of two polyketide chains joined through an ester linkage. One chain is a nonaketide that undergoes cyclization to form an octahydronaphthalene ring system, and the other is a diketide, (2R)-2-methylbutyrate. LOVF synthesizes the diketide side chain of lovastatin by acylating Ser76 of Lov D following a ping-pong mechanism[37,38]. An inhibitor of 3-hydroxy-3-methylglutaryl-coenzyme A reductase catalyses the synthesis of alpha-S-methylbutyrate using malonyl-coenzyme A, S-adenosyl-L-methionine and NADPH.

We have identified PKS at locus MPH11434 in *M. phaseolina* clustered with *Cochliobolus heterostrophus* PKS 3 (CHPKS3) and having siblings like *A. terreus* Lov F, which synthesizes diketide portion of lovastatin, presumably acting as a precursor for lovastatin.

Melanin pigment: Natural pigments are non-toxic, non-polluting and less hazardous to health. Moreover, their antioxidant and antimicrobial nature adds to their positive effects. Natural pigments can be obtained from various sources like plants, animals, fungi and microbes. Microbial

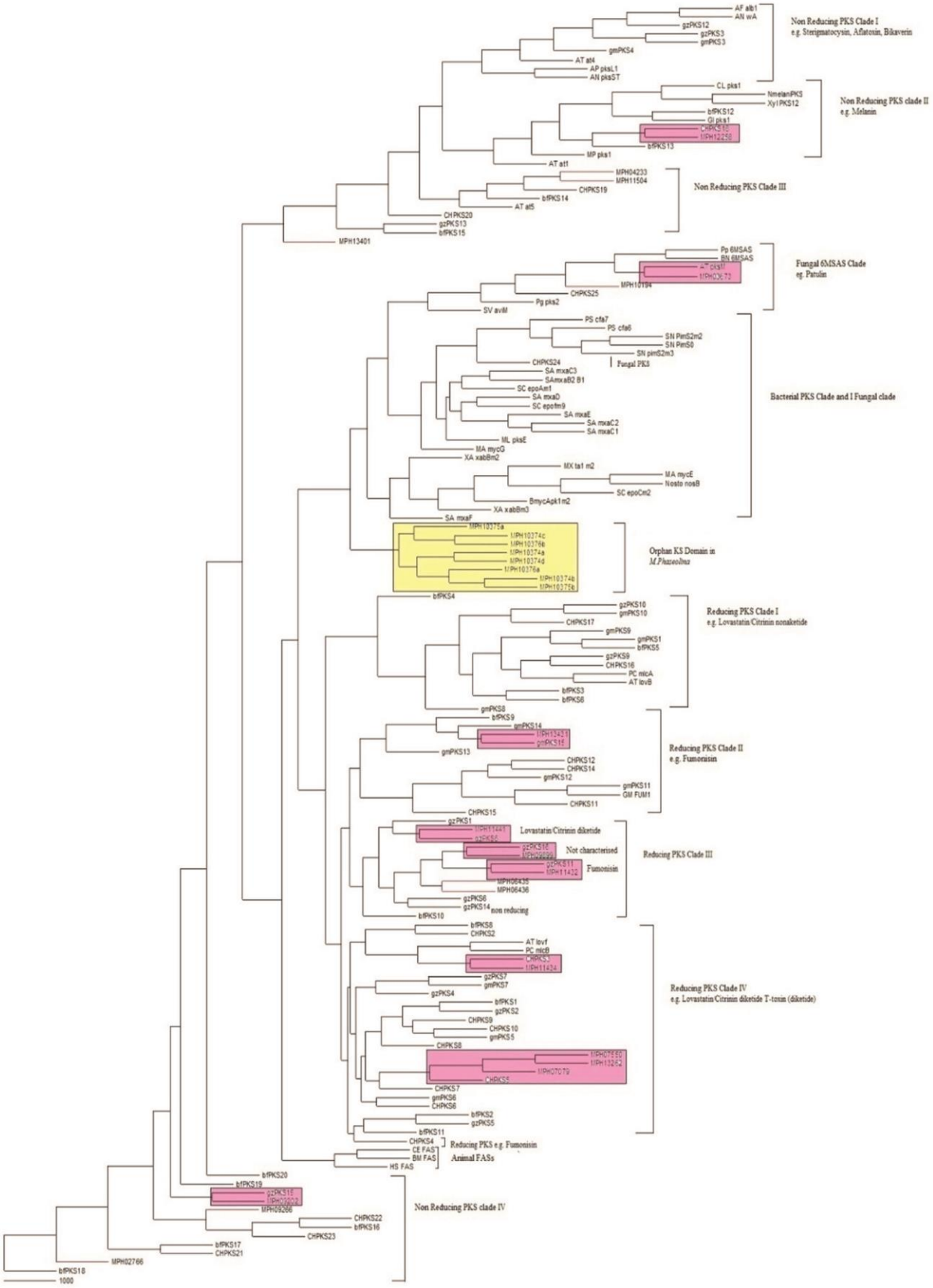

Figure 2. NJ phylogentic tree of KS domains of type-1 PKS of *Macrophomina phaseolina* with 120 KS domains of pathogenic saprobic ascomycecets fungi, bacteria and eukaryotes (obtained from Kroken *et al.*[32]) using Phylip program, setting the bootstrap analysis to 1000 runs.

pigments have some advantages over plant- and animal-based pigments, as microbes are fast-growing and have the potential of being standardized commercially. Attempts have been made to synthesize bacterial and fungal pigments to be used in the textile and leather industry. Many fungi like *Trichoderma virens*, *Curvularia lunata* and *Alternaria alternata* produce pigments, likewise the present study also predicts a polyketide that may encode melanin pigment[39]. Recently, genome sequence analysis of industrial strain *Monascus purpureus* has revealed the pigment biosynthesis pathway.

The present study shows that PKS of *M. phaseolina* falls under non-reducing clade 2. In this clade PKS at locus MPH12258 clusters with *Cochliobolus heterostrophus* PKS18 (CHPKS18), which is responsible for melanin pigment synthesis.

Mycotoxins: These are secondary metabolites of fungi that are harmful to both animals and humans. Mycotoxicoses are diseases caused by the ingestion of foods or feeds made toxic by these fungal metabolites. Therefore, these have importance in agricultural industry.

(i) Citrinin: This is a polyketide mycotoxin 4,6-dihydro-8-hydroxy-3,4,5-trimethyl-6-oxo-3H-2-benzopyran-7-carboxylic acid[40]. It is phenol derivative nephrotoxic mycotoxin produced by several fungal strains belonging to the genera *Penicillium*, *Aspergillus* and *Monascus*[41]. It contaminates various commodities of plant origin, cereals in particular. In addition to nephrotoxicity, citrinin is also embryocidal and fetotoxic. Clinically, citrinin is shown to cause renal disease in poultry, pigs, dogs and rats. The electron transport system of the kidney and liver mitochondria was considered as the target of the toxic action of citrinin. Citrinin contamination of food and feed is rather scarce. However, it is reasonable to believe that humans are much more frequently exposed to citrinin than generally accepted, because it is produced by the same moulds as ochratoxin A, which is a common contaminant of human food all over the world[42]. It also acts as a precursor for lovastatin diketide, which is a cholesterol-lowering agent[43].

The present study reveals that PKS at locus MPH11434 in *M. phaseolina* clusters with *C. heterostrophus* PKS 3 (CHPKS3) and has sibling *Penicillium* MLC B, which synthesizes diketide portion of citrinin, presumably acting as a precursor for citrinin.

(ii) Fumonisin: Fumonisins are naturally occurring toxins produced by the moulds *Fusarium moniliforme* (*F. verticillioides*), *F. proliferatum* and other *Fusarium* species. Among the identified fumonisins produced by the fungus *F. verticillioides*, B1, B2 and B3 are the most abundant in contaminated food and feed. Fumonisin B1 is the most common, comprising approximately 75% of the total fumonisin content, which occurs mainly in maize (corn), wheat and other cereals. Fumonisins can contaminate grain used for human food or livestock feed, as well as silage. Infection is increased if kernels are physically damaged, especially by insect feeding. Fumonisins can be found in a few other crops, typically at low levels, but their economic importance is mainly in maize. This mycotoxin has been associated with equine leukoencephalomalacia and pulmonary oedema in swine. It has also been associated with esophageal cancer in humans in certain regions of South Africa and China, where consumption of fumonisin-contaminated corn has been common. In addition to their adverse effect on the brain, liver and lungs, fumonisins also affect the kidneys, pancreas, testes, thymus, gastrointestinal tract and blood cells. There is also concern that consumption of fumonisins during early pregnancy could result in elevated risk of neural tube defect in the developing foetus. Fumonisin Bl works by inhibiting the biosynthesis of sphingosine and increasing the level of sphinganine, a sphingosine precursor in rat. Thus, fumonisin inhibition of sphingolipid metabolism could be at the core of a wide range of physiological effects, including tumour promotion. Fumonisins pose a potential health threat to both animals and humans, and these compounds cause several animal diseases. However, limited surveys indicate that human consumers of corn-based foods may be ingesting fumonisins[44].

The present study shows that PKS at locus MPH11432 has been grouped with *Gibberella zea* PKS 11 (GZPKS11) in reducing clade 3, which encodes product fumonisin.

(iii) Patulin: This is a toxic fungal metabolite produced by certain moulds of the genera *Penicillium*, *Aspergillus* and *Byssochlamys* growing on various food commodities, especially fruits. Patulin exhibits a number of toxic effects in animals, and its presence in food is undesirable. Patulin is associated with fresh fruits like apples, pears, apricots, peaches and grapes. Patulin is a relatively uncomplex lactone, ($C_7H_6O_4$), which is highly toxic to plant and animal cells. It can react with the terminal sulfhydryl groups of proteins and polypeptides present in food. It is found to inhibit protein and RNA synthesis, and is shown to produce persistent breaks in single and double strands of DNA in *Escherichia coli*. It can cause oxidative damage to DNA in human cells and also inhibit the activity of numerous enzymes, mainly as a consequence of its strong affinity to sulfhydryl groups. It has an inhibitory effect on several biochemical parameters such as ATPase, alkaline phosphatase, aldolase and hexokinase activity. Exposure of humans to patulin via consumption of infected products may result in severe toxicosis, including mutagenic, teratogenic, hepatotoxic, nephrotoxic, neurotoxic and genotoxic effects; its acute effects include nausea, vomiting and other gastrointestinal traumas that accompany kidney damage[45].

The genome of *M. phaseolina* has PKS at locus MPH13431 grouped with *Gibberella moniliformis* PKS

15 (GMPKS15) and placed in fungal 6-MSAS-type clade that is responsible for producing patulin.

Therefore, in order to initiate the activation of silent fungal secondary metabolite gene clusters, a number of strategies are developed (use of inducer, co-cultivation with bacteria, epigenetic re-modelling, use of rare earth elements, ribosomal engineering and others)[46–48]. Further experimental strategies may be applied to these existing strategies to validate if these aforementioned PKS are responsible for producing these potentially important natural products. In addition, we have identified eight KS domains from three reducing PKS, i.e. MPH10374, MPH10375 and MPH10376. These are called orphan polyketides because they do not have any known detectable homologs with fungi, bacteria and other organisms. The orphan polyketide may produce large novel new molecules that are yet to be explored.

Conclusion

The present study revealed that genome of charcoal rot fungi *M. phaseolina* contains several PKS, i.e. non-reducing, partially reducing and reducing types. Phylogenetic analysis using amino acid sequence of KS revealed differences among them. We also have analysed the position of *M. phaseolina* PKS in the phylogenetic tree by examining it with other PKS and fatty acid synthase obtained from the genome of other fungi, bacteria and eukaryotes used by Kroken *et al.*[32]. Until now it was unclear what chemical defence makes this fungus more fit in different plant-specific hosts and ecological niches. The diversity in the chemical structures of polyketides may have evolved as a result of continuous evolutionary pressure that resulted in chemical novelty facilitated by the modular nature of PKS. On an evolutionary scale, diversity is introduced into polyketides through gene flow, genetic recombination and mutation, i.e. both simple mutations within the domains and frequent horizontal co-transfer of genes between clusters[49]. Generation of novelty through exchange of domains between biosynthetic gene clusters polished under evolutionary selection pressure, invariably results in successful product assembly – as millions of failed experiments were rapidly discarded by natural selection. It is hypothesized that diversification of polyketides can occur in four steps throughout biosynthesis resulting from: (i) choice of polyketide building blocks and chain length; (ii) the extent of reduction and stereochemistry of β-keto intermediates primary cyclization, alkylation and branching; (iii) rearrangements and secondary cyclization, and (iv) post-polyketide tailoring: glycosylation and oxygenation[50,51]. Therefore, the presence of a large number of polyketides and orphan polyketides indicates that *M. phaseolina* may possess sophisticated genetic mechanisms that facilitate its adaptation to heterogenous environments such as soil and living plant host. Further analysis has allowed us to predict polyketide-natural products of *M. phaseolina*, which have close similarity with lovastatin, mycotoxins and pigment. The present study hints at the immense potential of *M. phaseolina* to provide industrially useful and life-saving drugs, dye and toxins. Further research is required for its validation and proper prospecting.

1. Brakhage, A. A., Schuemann, J., Bergmann, S., Scherlach, K., Schroeckh, V. and Hertweck, C., Activation of fungal silent gene clusters: a new avenue to drug discovery. *Prog. Drug Res.*, 2008, **66**, 3–12.
2. Gomes, E. S., Schuch, V. and MacedoLemos, E. G., Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz. J. Microbiol.*, 2014, **44**, 1007–1034.
3. Helfrich, E. J., Reiter, S. and Piel, J., Recent advances in genome-based polyketide discovery. *Curr. Opin. Biotechnol*., 2014, **29**, 107–115.
4. Hoffmeister, D. and Keller, N. P., Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat. Prod. Rep.*, 2007, **24**, 393–416.
5. Newman, D. J. and Cragg, G. M., Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.*, 2007, **70**, 461–477.
6. García-Cela, E., Crespo-Sempere, A., Gil-Serna, J., Porqueres, A. and Marin, S., Fungal diversity, incidence and mycotoxin contamination in grapes from two agro-climatic Spanish regions with emphasis on *Aspergillus* species. *J. Sci. Food. Agric.*, 2015, **95**, 1716–1729.
7. Yang, Y. *et al.*, Complete genome sequence and transcriptomics analyses reveal pigment biosynthesis and regulatory mechanisms in an industrial strain, *Monascus purpureus* YY-1. *Sci. Rep.*, 2015, **5**, 8331.
8. Brakhage, A. A., Bergmann, S., Schuemann, J., Scherlach, K., Schroeckh, V. and Hertweck C., Fungal genome mining and activation of silent gene clusters. In *The Mycota XV* (ed. Esser, K.), Springer-Verlag, Heidelberg, Germany, 2009, pp. 297–303.
9. Chiang, Y. M., Lee, K. H., Sanchez, J. F., Keller, N. P. and Wang, C. C., Unlocking fungal cryptic natural products. *Nat. Prod. Commun*., 2009, **11**, 1505–1510.
10. Fedorova, N. D., Moktali, V. and Medema, M. H., Bioinformatics approaches and software for detection of secondary metabolic gene clusters. In *Fungal Secondary Metabolism: Methods and Protocols* (eds Keller, N. P. and Turner, G.), Springer, Germany, 2012, pp. 23–45.
11. Liu, L., Zhang, Z., Shao, C. L., Wang, J. L., Bai, H. and Wang, C. Y., Bioinformatical analysis of the sequences, structures and functions of fungal polyketide synthase product template domains. *Sci. Rep*., 2015, **5**, 10463.
12. Jez, J. M., Austin, M. B., Ferrer, J., Bowman, M. E., Schröder, J. and Noel, J. P., Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.*, 2000, **7**, 919–930.
13. Dhingra, O. D. and Sinclair, J. B., Biology and pathology of *Macrophomina phaseolina*. Imprensa Universitaria, Universidade Federal de Vicosa, Vicosa, Brasil, 1978, p. 166.
14. Farr, D. F., Bills, G. F., Chamuris, G. P. and Rossman, A. Y., *Fungi on Plants and Plant Products in the United States*, APS Press, 1989, pp. 405–407.
15. Mihail, J. D. and Taylor, S. J., Interpreting variability among isolates of *Macrophomina phaseolina* in pathogenicity, pycnidium production and chlorate utilization. *Can. J. Bot.*, 1995, **73**, 1596–1603.

16. Islam, M. S. *et al.*, Tools to kill: genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics*, 2012, **13**, 493.
17. Wrather, J. A., Koenning, S. R. and Anderson, T. R., Effect of diseases on soybean yields in the United States and Ontario (1999–2002). *Plant Health Prog.*, 2003 (online).
18. Kaur, S., Dhillon, G. S., Brar, S. K., Vallad, G. E., Chand, R. and Chauhan, V. B., Emerging phytopathogen *Macrophomina phaseolina*: biology, economic importance and current diagnostic trends. *Crit. Rev. Microbiol.*, 2012, **38**, 136–151.
19. Cacho, R. A., Tang, Y. and Chooi, Y. H., Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Front. Microbiol.*, 2015, **5**, 774.
20. Challis, G. L., Genome mining for novel natural product discovery. *J. Med. Chem.*, 2008, **51**, 2618–2628.
21. Delgado, J. A., Al-Azzam, O., Denton, A. M., Markell, S. G. and Goswami, R. S., A resource for the *in silico* identification of fungal polyketide synthases from predicted fungal proteomes. *Mol. Plant Pathol.*, 2012, **13**, 494–507.
22. Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A. and Birren, B., Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.*, 2005, **15**, 1620–1631.
23. Guo, C. J. and Wang, C. C., Recent advances in genome mining of secondary metabolites in *Aspergillus terreus*. *Front. Microbiol.*, 2014, **5**, 717.
24. Medema, M. H. *et al.*, Minimum information about a biosynthetic gene cluster. *Nature Chem. Biol.*, 2015, **11**, 625–631.
25. Medema, M. H. and Fischbach, M. A., Computational approaches to natural product discovery. *Nature Chem. Biol.*, 2015, **11**, 639–648.
26. Hadjithomas, M. *et al.*, IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *mBio*, 2015, **6**, 00932–01015.
27. O'Brien, R. V., Davis, R. W., Khosla, C. and Hillenmeyer, M. E., Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiot.*, 2014, **67**, 89–97.
28. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. and Borodovsky, M., Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.*, 2008, **18**, 1979–1990.
29. Weber, T., *In silico* tools for the analysis of antibiotic biosynthetic pathways. *Int. J. Med. Microbiol.*, 2014, **304**, 230–235.
30. Medema, M. H. *et al.*, anti-SMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, 2011, **39**, W339–W346.
31. Blin, K., Medema, H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E. and Weber, T., antiSMASH 2.0 – a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.*, 2013, **41**, W204–W212.
32. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. and Turgeon, B. G., Phylogenomic analysis of type-I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 15670–156755.
33. Thompson, J. D., Gibson, T. J. and Higgins, D. G., Multiple sequence alignment using Clustal W and Clustal X. In *Curr. Protoc. Bioinformatics*, 2002, 2:2.3.1–2:2.3.22.
34. Baum, B. R., PHYLIP: Phylogeny Inference Package. Version 3.2 (Software review). *Quart. Rev. of Biol.*, 1989, **64**, 539–541.
35. Felsenstein, J., PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle, USA, 1993.
36. Mansour, A. (ed.), Phylip and phylogenetics. In *Focus on Bioinformatics: Genes, Genomes and Genomics*, 2009, vol. 3, pp. 46–49.
37. Xie, X., Meehan, M. J., Xu, W., Dorrestein, P. C. and Tang, Y., Acyltransferase mediated polyketide release from a fungal megasynthase. *J. Am. Chem. Soc.*, 2009, **131**, 8388–8389.
38. Osuna, S., Jiménez-Osés, G., Noey, E. L. and Houk, K. N., Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Chem. Res.*, 2015, **48**, 1080–1089.
39. Sharma, D., Gupta, C., Aggarwal, S. and Nagpal, N., Pigment extraction from fungus for textile dyeing. *Indian J. Fibre Text. Res.*, 2012, **37**, 68–73.
40. Clark, B. R., Capon, R. J., Lacey, E., Tennant, S. and Gill, H., Citrinin revisited: from monomers to dimers and beyond. *Org. Biomol. Chem.*, 2006, **4**, 1520–1528.
41. Xu, B., Jia, X., Gu, L. and Sung, C., Review on the qualitative and quantitative analysis of the mycotoxin citrinin. *Food Control*, 2004, **17**, 271–285.
42. Martin, D. R., Scientific opinion on the risks for public and animal health related to the presence of citrinin in food and feed. EFSA Panel on Contaminants in the Food Chain (CONTAM). European Food Saf. Authority, Parma, Italy, 2012, vol. 10, p. 2567.
43. Endo, A. and Kuroda, M., Citrinin, an inhibitor of cholesterol synthesis. *J. Antibiot.*, 1976, **19**, 841–843.
44. Nelson, P. E., *Fumonisins, mycotoxins* produced by *Fusarium* species: biology, chemistry, and significance. *Annu. Rev. Phytopathol.*, 1993, **31**, 233–252.
45. Puel, O., Galtier, P. and Oswald, I. P., Biosynthesis and toxicological effects of patulin. *Toxins*, 2010, **2**, 613–631.
46. Brakhage, A. A. and Schroeckh, V., Fungal secondary metabolites – strategies to activate silent gene clusters. *Fungal Genet. Biol.*, 2011, **48**, 15–22.
47. Okhi, K. and Hosaka, T., New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl. Microbiol. Biotechnol.*, 2013, **97**, 87–98.
48. Netzker, T. *et al.*, Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front. Microbiol.*, 2015, **20**, 299.
49. Mehrabi, R. *et al.*, Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. *FEMS Microbiol. Rev.*, 2012, **35**, 542–554.
50. Hertweck, C., Hidden biosynthetic treasures brought to light. *Nat. Chem. Biol.*, 2009, **7**, 450–452.
51. Chooi, Y. H. and Tang, Y., Navigating the fungal polyketide chemical space: from genes to molecules. *J. Org. Chem.*, 2012, **77**, 9933–9953.

Received 19 February 2016; revised accepted 8 August 2016

doi: 10.18520/cs/v112/i02/322-331