

# SCIENTIFIC REPORTS

OPEN

## Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains

Received: 25 November 2014

Accepted: 15 April 2015

Published: 21 May 2015

Lu Liu<sup>1</sup>, Zheng Zhang<sup>2</sup>, Chang-Lun Shao<sup>1</sup>, Jin-Lan Wang<sup>3</sup>, Hong Bai<sup>1</sup> & Chang-Yun Wang<sup>1,4</sup>

The product template (PT) domains, specifically in fungal non-reducing polyketide synthases (NR-PKSs), mediate the regioselective cyclization of polyketides dominating the final structures. However, up to date, the systematic knowledge about PT domains has been insufficient. In present study, the relationships between sequences, structures and functions of the PT domains were analyzed with 661 NR-PKS sequences. Based on the phylogenetic analysis, the PT domains were classified into prominent eight groups (I–VIII) corresponding with the representative compounds and cyclization regioselectivity (C<sub>2</sub>-C<sub>7</sub>, C<sub>4</sub>-C<sub>9</sub>, and C<sub>6</sub>-C<sub>11</sub>). Most of the cavity lining residue (CLR) sites in all groups were common, while the regional CLR site mutations resulted in the appearance of finger-like regions with different orientation. The cavity volumes and shapes, even the catalytic dyad positions of PT domains in different groups were corresponding with characteristic cyclization regioselectivity and compound sizes. The conservative residues in PT sequences were responsible for the cyclization functions and the evolution of the key residues resulted in the differentiations of cyclization functions. The above findings may help to better understand the cyclization mechanisms of PT domains and even predict the structural types of the aromatic polyketide products.

One hallmark of fungi is their capacity to synthesize diverse biological polyketide natural products with structural variation during synthesis by polyketide synthase (PKSs)<sup>1–3</sup>. With a few exceptions, a majority of fungal PKSs are iterative type I PKSs<sup>1,4</sup>. Fungal PKSs are intrinsically more difficult to be studied than the dissociated bacterial PKSs due to their large sizes (in excess of 200 kDa) and the difficulties with genetic manipulation<sup>5</sup>. In the past decade, more knowledge about biosynthesis mechanisms of fungal polyketide synthases has been acquired. It is generally recognized that most biosynthetic gene clusters including PKSs are silent or expressed at very low levels according to global transcriptomic profile<sup>6</sup>. Fungi have the potential to produce a far greater number of polyketides than the known polyketides isolated from fungi up to date. Therefore, it is essential to investigate and characterize fungal PKSs for enzymatic mechanism elucidation and genetic manipulation to obtain more new metabolites.

The fungal PKSs can be divided into three major classes according to the function and phylogeny, i.e., the non-reducing (NR), the highly-reducing (HR), and the partial-reducing (PR) PKSs<sup>1</sup>. The NR-PKSs synthesize aromatic polyketides, such as carcinogenic mycotoxin aflatoxin<sup>7</sup>. Besides the basic

<sup>1</sup>Key Laboratory of Marine Drugs, the Ministry of Education of China, School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China. <sup>2</sup>State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, China. <sup>3</sup>School of Life Science, Shandong University, Jinan 250100, China. <sup>4</sup>Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China. Correspondence and requests for materials should be addressed to Z.Z. (email: zhzhang.sdu@gmail.com) or C.Y.W. (email: changyun@ouc.edu.cn)

PKS domains, additional functional domains including the starter unit-ACP transacylase (SAT) domain, product template (PT) domain, and thioesterase (TE) releasing domain are unique to the NR-PKSs<sup>1,7,8</sup>. Recently, the PT domains have been demonstrated to be embedded in controlling specific aldol cyclization and aromatization of the polyketide precursors<sup>7</sup>. There are three commonly cyclizing patterns (C2-C7, C4-C9, and C6-C11), when the first cyclization occurs<sup>9</sup>.

According to an earlier phylogenetic analysis of  $\beta$ -ketoacyl synthase (KS) domains, NR-PKSs were divided into three basal subclades (subclades I–III), each of them characterized a typical domain architecture<sup>10</sup>. Subclade III was characterized by a methyltransferase (CMeT) domain located after the acyl carrier protein (ACP) domain. And subclades I and II diverged after losing the CMeT domain. Later, based on a phylogenetic analysis of PT domains in subclades I and II from known functional NR-PKSs, the PT domains were classified into five groups (groups I–V) corresponding to the cyclization regioselectivity at C2-C7, C4-C9, and C6-C11 as well as compound sizes<sup>9</sup>. Recently, the PT domains with associated metabolites in subclade III were further classified into two groups (groups VI and VII) responsible to produce aromatics with C2-C7 cyclization<sup>11</sup>.

Up to date, only one PT domain (group IV) crystal structure has been studied from *Aspergillus parasiticus* PksA, which participates in aflatoxin B1 biosynthesis<sup>12</sup>. The PT structure of PksA displays a distinct ‘double hot dog’ (DHD) fold and the two DHD monomers associate *via* a PT-specific sequence insertion. The residues important for substrate binding and catalysis were revealed, and an internal pocket was identified, which can be visually divided into three regions, phosphopantetheine (PPT)-binding region, cyclization chamber, and hydrophobic hexyl-binding region. The crystal structure of PT domain in PksA provided a vital model for the structural study on other PT domains of which structures were still unknown.

So far, the systematic knowledge about PT domains has been insufficient. The relationships between sequences, structures and functions of the PT domains have been unclear. In this study, the NR-PKS phylogenetic tree and PT phylogenetic tree were established based on all known fungal NR-PKSs protein sequences. The three-dimensional structures of different PT domains were modeled, and the relationships between catalytic pocket shapes and sizes, regioselective cyclization, and compound sizes were analyzed. The influence of the PT domain sequence variations on the differentiations of structures and functions was also discussed.

## Results and Discussion

**Phylogenetic analysis of fungal NR-PKSs.** All of the known fungal NR-PKSs amino acid sequences were searched and screened from the NCBI database to establish phylogenetic trees (Supplementary Table S1). The collected NR-PKSs dataset comprised of 661 sequences, of which 627 sequences were derived from 187 strains of ascomycetes and 34 sequences from 29 strains of basidiomycetes.

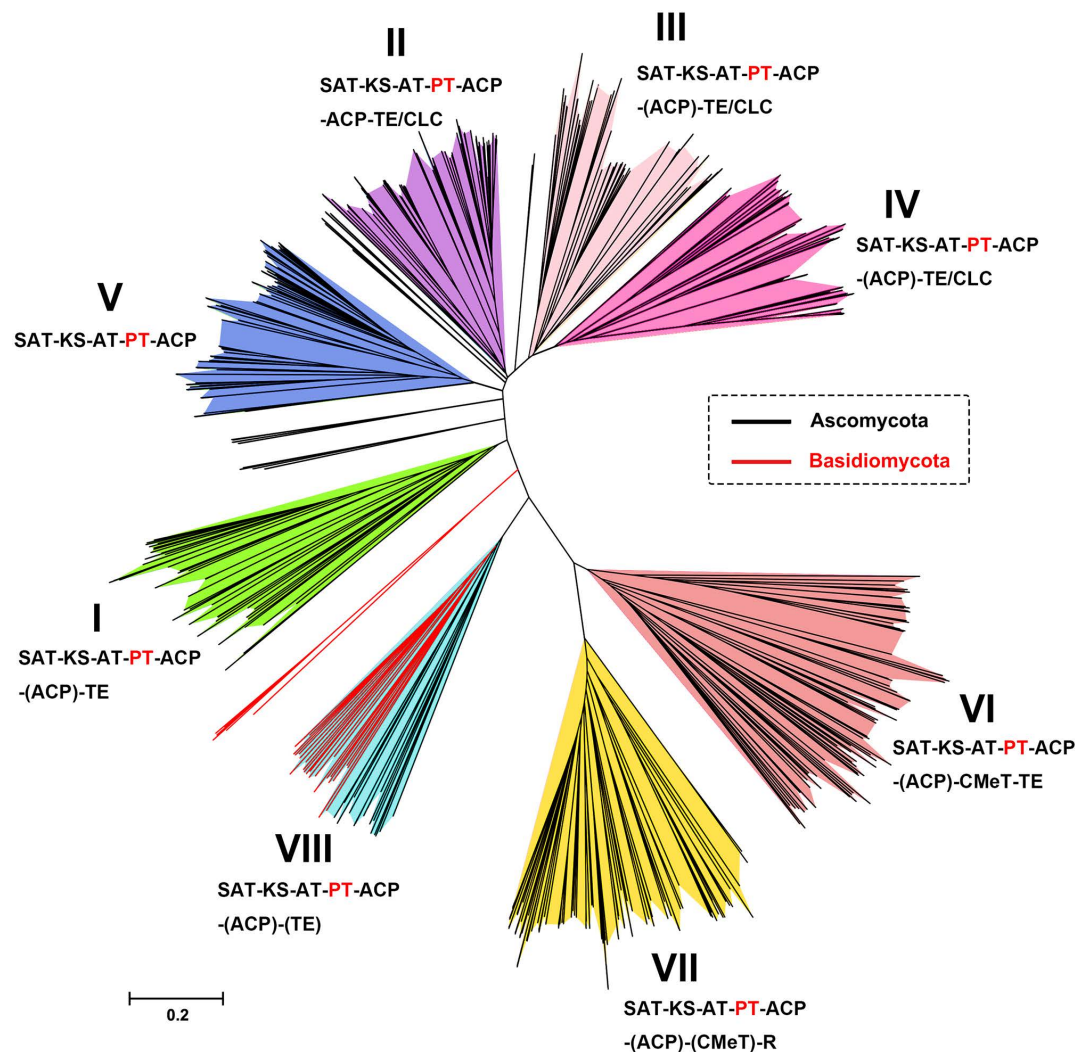
The NR-PKS phylogenetic tree was constructed on the basis of the phylogeny and domain architectures. The resulting NR-PKS phylogenetic tree clearly classified the collected NR-PKS sequences into eight major groups (groups I–VIII) except a few sequences (Fig. 1). Compared with the previous report<sup>11</sup>, besides the seven groups, a new group (VIII) appeared in this tree obviously different from the other groups. The NR-PKSs from ascomycetes were found to cover throughout all eight groups, however the NR-PKSs from basidiomycetes only appeared in group VIII and rare intermediate clades. It was found that the NR-PKSs from each genus fell into no more than six groups. The NR-PKSs from genera *Aspergillus*, *Bipolaris*, *Colletotrichum* and *Talaromyces* were found in six groups in the phylogenetic tree.

All of the NR-PKSs contain SAT, KS, AT, PT and ACP domains with the PT domains locating between AT and ACP domains. Moderate differences of domain architectures were observed in eight groups. The C-terminus TE/CLC domains responsible for releasing products in all members in group V were absent. Most members in groups VI and VII were characterized by a CMeT domain located after the ACP domain, which can produce aromatics with methyl branching on their benzene rings. Most NR-PKSs (groups I–IV, VI, VIII) showed an additional ACP domain. A previous study revealed that the product-releasing domains including TE/CLC in each group could be further distinguished by different functions<sup>13</sup>. It suggests that more information about NR-PKS sequences should be accumulated to clarify the phylogenetic relationships between different groups in the future.

**Differentiation of PT domain functions.** The phylogenetic tree of PT domains from 661 NR-PKS sequences was also constructed. The PT sequence boundaries were defined by computation tools according to the NR-PKSs differentiation and the sequence length is about 300 ~ 400 amino acids. The established PT domain phylogenetic tree was similar to that of NR-PKSs, also classified into eight groups (Supplementary Fig. S1).

Until now, 55 NR-PKSs among the 661 NR-PKS sequences have been investigated for their regioselective cyclization modes (Supplementary Table S2). Successively, the PT domain phylogenetic tree from these 55 NR-PKSs was established based on their regioselective cyclization modes (Fig. 2). These 55 PT domains were also found to cover all eight groups (groups I–VIII), responding to the common first-ring cyclization of C2-C7 (groups I–III, VI–VIII), C4-C9 (group IV) and C6-C11 (group V), and the rare cyclization of C1-C6 (group V) and C3-C8 (group I).

The phylogenetic analysis suggested that each group of PT domains associates with specific regioselective cyclization and compound size. In group I, most PT domains catalyze the first-ring cyclization



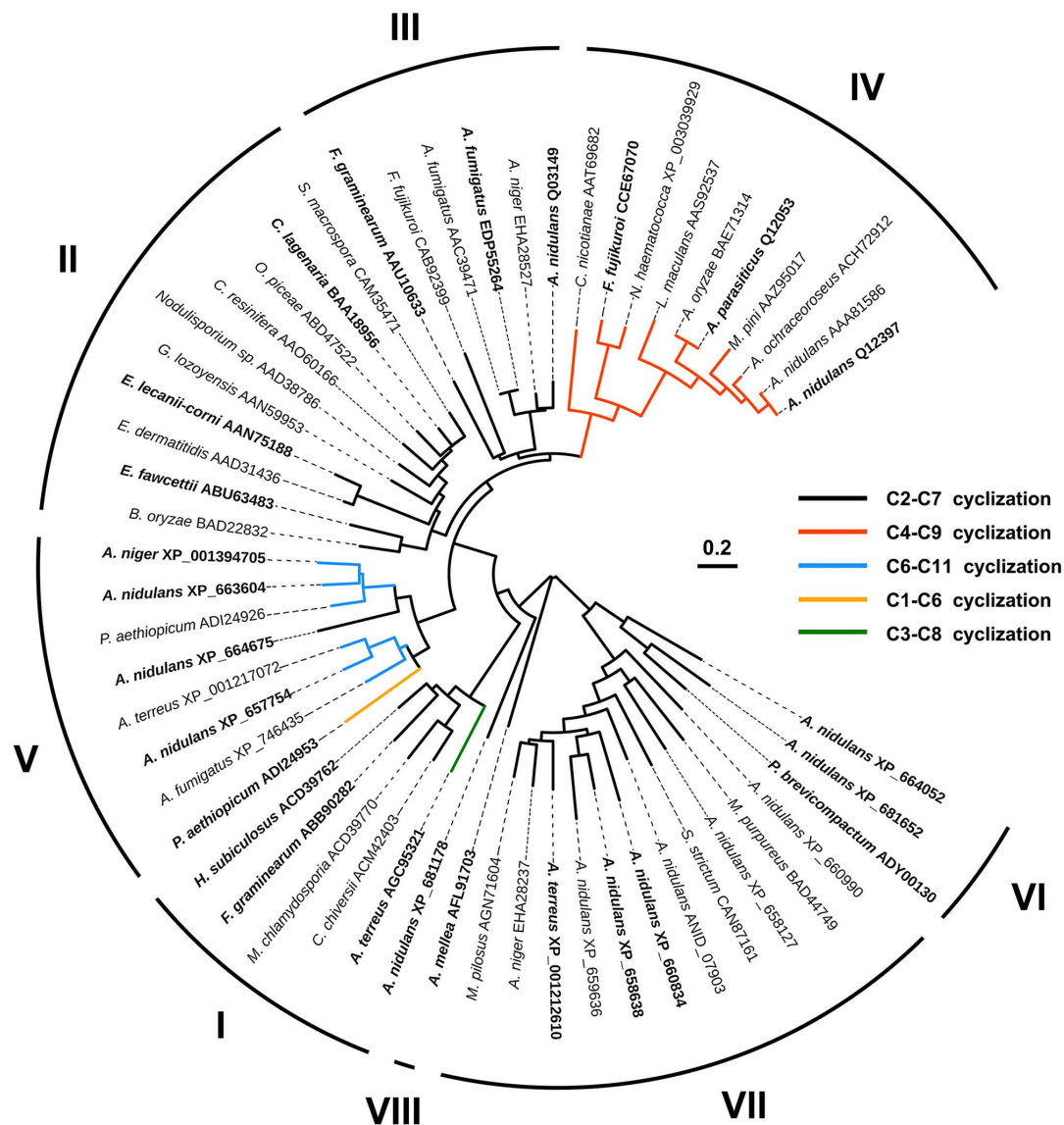
**Figure 1.** NR-PKS phylogenetic tree of 661 sequences from ascomycetes and basidiomycetes. The branches of eight groups have been colored, each of which shares a common organization of domains (those in parentheses are variable in their presence/absence within that group). Domain abbreviations: SAT = starter unit-ACP transacylase, KS = ketosynthase, AT = acyl transferase, PT = product template, ACP = acyl carrier protein, TE = thioesterase, TE/CLC = thioesterase/Claisen cyclase, CMeT = C-methyltransferase, R = reductase. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

via C2-C7 aldol condensation. The main NR-PKSs in this group are involved in the biosynthesis of aromatic portions in the resorcylic acid lactones, exemplified by hypothemycin Hpm3<sup>14</sup>, radicicol RDC1<sup>14</sup>/RADS2<sup>15</sup>, and zearalenone PKS13<sup>16</sup>. Group I also includes a single aromatic ring synthase, the orsellinic acid OrsA from *Aspergillus nidulans*<sup>17</sup>. As an exception, the dehydrocurvularin AtCURS2 from *Aspergillus terreus* forms the first rings via C3-C8 rather than C2-C7<sup>18</sup>.

In group II, the NR-PKSs include the most known THN synthases from a variety of genera responsible for the synthesis of 1,3,6,8-tetrahydroxynaphthalene (THN) analogues<sup>19</sup>. The PT domains of this group catalyze the aldol condensation via C2-C7 with pentaketide backbones. The TE/CLC domains in this group are in charge of cyclizing the second rings.

The NR-PKSs in group III synthesize longer polyketide chains via C2-C7 cyclization, such as naphthopyrone heptaketide synthase Alb1<sup>20</sup> and bikaverin nonaketide synthase PKS4<sup>21</sup>. Interestingly, the WA synthase from *A. nidulans* also generates pentaketide THN with shorter polyketide chains<sup>11</sup>. The previously biochemical analysis showed that THN is a result of shortening of the heptaketide YWA1, a common metabolite in group III synthesized by WA synthase<sup>11,22</sup>.

All the PT domains in group IV cyclize the first ring via C4-C9 regioselectivity. The only one crystal structure of PT domain in PksA from *A. parasiticus*<sup>12</sup> is classified into group IV. The NR-PKSs of this group contain the analogues of decaaketide synthase PksA, such as aflatoxin/sterigmatocystin

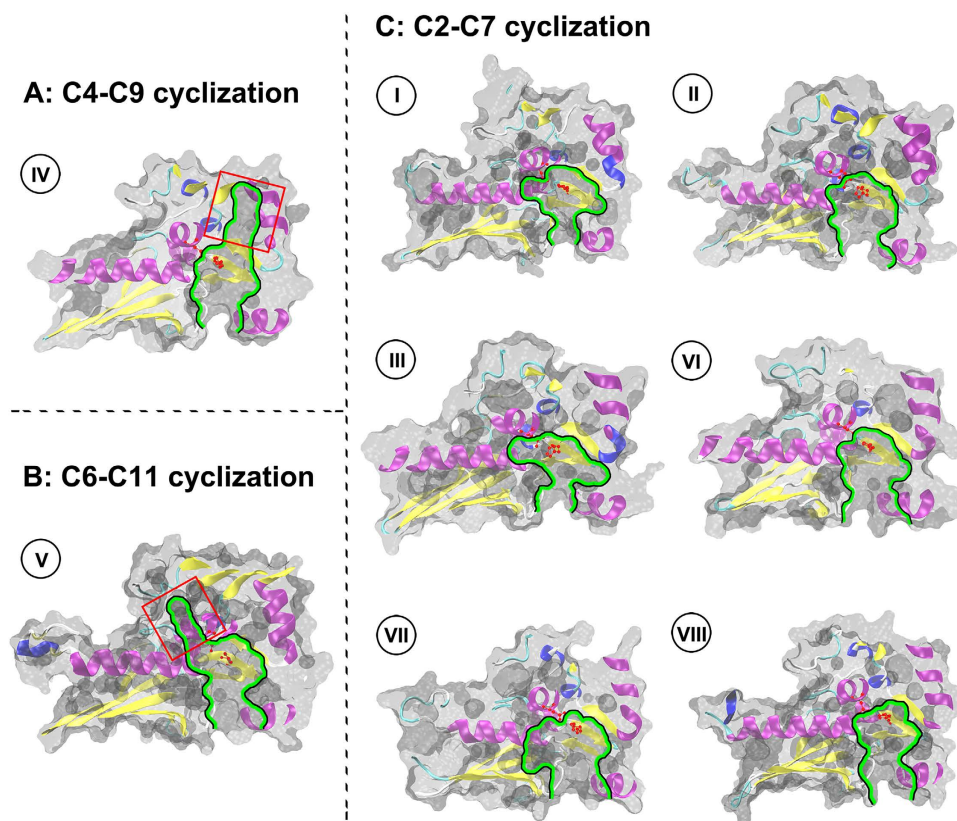


**Figure 2. Phylogenetic tree of PT domains of selected 55 NR-PKSs.** PT domains of 55 NR-PKSs are related to cyclization regioselectivities. Phylogenetic analysis was conducted using the bootstrap neighbor joining method. Bold branches indicate sequences which have been used in structure modeling. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

syntheses<sup>23,24</sup>. The NR-PKSs in this group also include heptaketide synthases such as cercosporin CTB1<sup>25</sup> and naphthaldehyde PKS1<sup>26</sup>.

All of the NR-PKSs in group V lack TE/CLC domains and they mainly form the first ring *via* C6-C11 register. It seems that these NR-PKS members produce multiple aromatic polyketides with the most different polyketide backbones by regioselectivity cyclization modes. Group V is composed of the NR-PKSs catalyzing the synthesis of different polyketide backbones, such as griseofulvin heptaketide synthase GsfA from *Penicillium aethiopicum*<sup>27</sup>, emodin octaketide synthase ACAS from *A. terreus*<sup>28</sup>, asperthecin nonaketide synthase AptA from *A. nidulans*<sup>29</sup> and TAN-1612 decaketide synthase AdaA from *Aspergillus niger*<sup>30</sup>. It should be noted that heptaketide synthase GsfA<sup>27</sup> catalyze the unorthodox C1-C6 aldol cyclization. Interestingly, heptaketide synthase PkgA<sup>11</sup> for C2-C7 aldol cyclization was also found in this group.

All of the NR-PKSs in groups VI and VII associate with C2-C7 specific cyclization, and most of them have a CMeT domain. The compounds derived from group VI are single aromatic ring natural products such as 5-methylorsellinic acid from *Penicillium brevicompactum*<sup>31</sup>. While the metabolites synthesized by the NR-PKSs of group VII were found to have different polyketide backbones, such as tetraketide orsellinaldehyde<sup>11</sup>, pentaketide citrinin<sup>32</sup>, hexaketide rubropunctatin<sup>33</sup>, octaketide



**Figure 3. Comparison of pocket shapes among eight groups of PT domains.** A. The PT structure for C4-C9 cyclization is exemplified by 3HRR PDB structure from group IV. B. The PT structure for C6-C11 cyclization is exemplified by XP\_657754 structural model from group V. C. The PT structure for C2-C7 cyclization is exemplified by XP\_681178 from group I, BAA18956 from group II, Q03149 from group III, XP\_681652 from group VI, XP\_658638 from group VII and AFL91703 from group VIII. All PT structures for C2-C7 cyclization are structural models. Surface representations of the pocket in the PT monomer are shown using green outline. The special finger-like regions of group IV and group V are respectively indicated with red boxes. The side chain atoms of catalytic dyads (His/Asp) in all groups are highlighted.

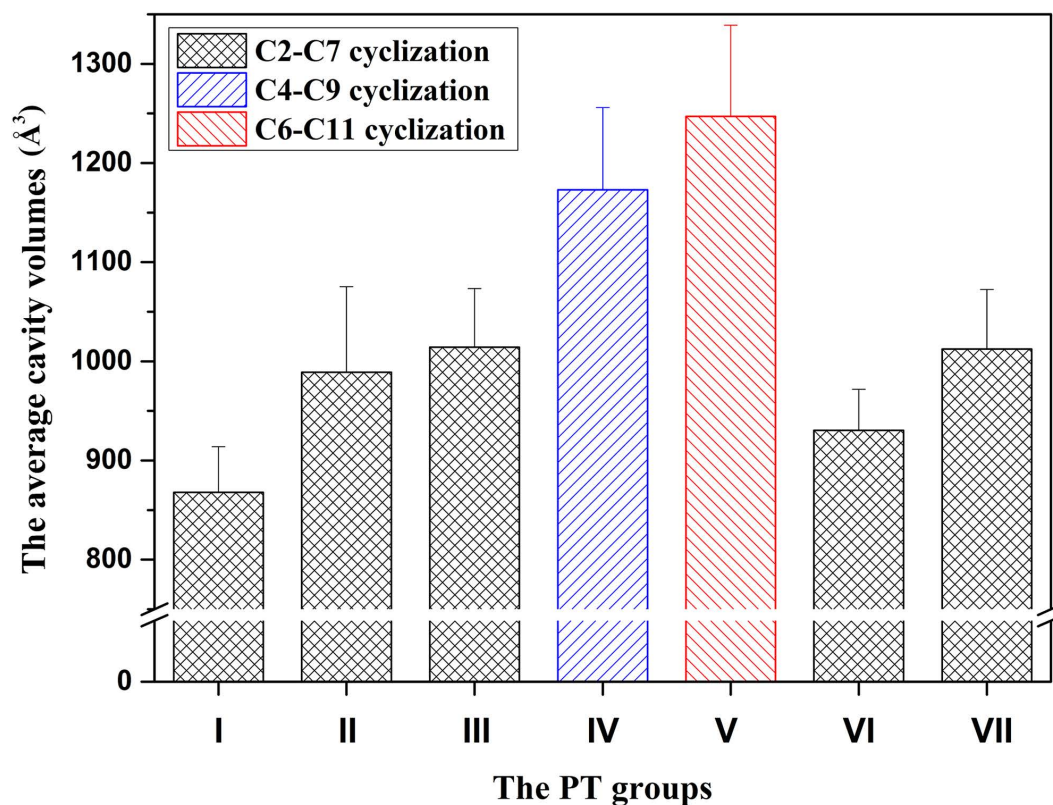
2,4-dihydroxy-6-[(3E,5E,7E)-2-oxonona-3,5,7-trienyl]benzaldehyde<sup>11</sup>, and nonaketide 2,4-dihydroxy-3-methyl-6-(2-oxopropyl)benzaldehyde<sup>11</sup>.

Group VIII is a newly separate group according to our phylogenetic analysis. Different from other groups, this group contains not only the NR-PKS members from ascomycetes and but also from basidiomycetes. The only studied polyketide synthase in group VIII is orsellinic acid synthase ArmB from *Armillaria mellea* with C2-C7 cyclization<sup>34</sup>.

The above phylogenetic analysis for PT domains indicated that the cyclization modes of C2-C7, C4-C9, C6-C11, C1-C6, and C3-C8 were found in eight groups. C2-C7 cyclization is the most common mode that spreads over seven groups except group IV. And the C4-C9 and C6-C11 cyclization modes only appear in groups IV and V, respectively.

**Structural analysis of PT domains.** In order to describe the PT domain structures corresponding to various cyclization modes, the catalytic pockets of different PT domains were analyzed and compared. The PT domain of *A. parasiticus* PksA responsible for C4-C9 cyclization has been the only PT crystal structure reported to date. The internal pocket of this PT domain extends 30 Å from the surface to the bottom (Fig. 3A). The catalytic pocket is divided into three regions: the phosphopantetheine (PPT)-binding region, the cyclization chamber, and the hydrophobic hexyl-binding region. The proposed catalytic dyad (Asp 1543/His 1345) locates at the middle of the pocket, i.e. the cyclization chamber, and initiates the regiospecific cyclizations appropriately.

The three-dimensional PT structures of eight groups were built by comparative protein modeling methods. For each group except for group VIII, at least three representative PT sequences were used to build three-dimensional structures (Supplementary Table S3). All of the sequences analyzed by comparative protein modeling methods were highlighted in bold in PT domain phylogenetic tree (Fig. 2). In order to make the simulated structures more convincing, three structural models for each PT domain were acquired by SWISS-MODEL<sup>35</sup>, I-TASSER<sup>36</sup>, and PHYRE2<sup>37</sup>, respectively. All the structural models



**Figure 4.** The average value of cavity volumes for PT groups (groups I–VII). The cavity volumes were calculated and identified by CASTp with 1.4 Angstroms probe radius.

were refined in the atomic-level by the fragment-guided molecular dynamics (FG-MD) simulations<sup>38</sup>. According to quality assessment by TM-score<sup>39</sup> and Ramachandran plot<sup>40</sup>, the accuracy of all the PT protein structural models were acceptable.

Based on the above simulated structures, the PT structures for C6-C11 cyclization were found to have a catalytic pocket deep about 30 Å from the surface to the bottom. Furthermore, the proposed catalytic dyad locates in the middle of the pocket. These characteristics of C6-C11 cyclization are similar to that of C4-C9 register (Fig. 3B). The obvious difference between C6-C11 and C4-C9 was found in the finger-like regions. Contrasted to the long, straight region of C4-C9, the finger-like region of C6-C11 orientates at a specific angle. In the PT structure for C2-C7, contrasting to that of C4-C9 and C6-C11, the depth of catalytic pockets are obviously shallow, range from 15 Å to 20 Å. It should be pointed out that C2-C7 catalytic pockets in groups I–III and VI–VIII are lack of the finger-like regions and the catalytic dyads locate at the bottom of the pockets (Fig. 3C).

The above results indicated that the features of PT pockets adapt with various catalytic mechanisms. In C4-C9 and C6-C11 cyclization modes, the catalytic pockets have larger sizes with the finger-like regions. The previous crystal structure study revealed that the hydrophobic hexyl-binding region (the finger-like region) of PT domain in aflatoxin synthase PksA can perfectly adapt to accept the substrate hexanoyl starter unit<sup>12</sup>. In present study, based on the three-dimensional model analysis, the PT domains of C4-C9 cyclization present the similar finger-like regions and generate aflatoxin analogues (Supplementary Table S2). While in the PT domains of C6-C11 cyclization, the finger-like regions associated with diverse polyketides was observed for the first time. It suggests that the finger-like regions for C4-C9 and C6-C11 cyclizations can accommodate the relevant chains of these compounds during the ring formation processes. Correspondingly, in the PT crystal structure of PksA, the proposed catalytic dyad appropriately locates in the middle of catalytic pockets<sup>12</sup>. Similarly, in the PT domains of C4-C9 and C6-C11, the proposed catalytic dyads also locate in the middle of catalytic pockets like PksA to initiate the regiospecific cyclizations. In all groups of C2-C7 PT catalytic pockets, the finger-regions have not been observed and the catalytic dyads locate at the bottom of the pockets. It suggests that these catalytic dyad positions are corresponding with the C2-C7 cyclization and the polyketide chains could extend or curve to close to the catalytic dyads during the ring cyclization processes.

In order to analyze the relationship between the cavity volumes of PT pockets and compound sizes, the cavity volumes of PT catalytic pockets were calculated by using CASTp (Supplementary Table S4). It was found that the average cavity volume of C6-C11 (group V) pockets is the biggest one, followed by

the C4-C9 (group IV) pockets (Fig. 4). The cavity volumes of C2-C7 pockets in five groups (groups I-III, VI, and VII) are significantly smaller than those of C6-C11 and C4-C9 pockets, with groups I, II and VI smaller than those of groups III and VII.

The C2-C7 cyclization modes cover the groups I-III, and VI-VIII. The NR-PKs in groups I, VI and VIII generate the tetraketides as the common metabolites. Interestingly, the compounds generated by groups I, II and VI, such as orsellinic acid<sup>17,31</sup> and THN<sup>19</sup>, are smaller than the compounds derived from groups III and VII, such as YWA1<sup>11</sup> and asperfuranone<sup>41</sup>. It seems that the slightly bigger volumes of PT catalytic pockets in groups III and VII make it possible for the longer chains to be accommodated and cyclized. In fact, the THN synthases appear in both groups II and III<sup>11,19</sup>, and most are involved in group II. As mentioned above, in group III, the THN is derived from heptaketide YWA1 synthesized by WA synthase<sup>22</sup>. The NR-PKs in group III usually synthesize compounds with longer chains and multiply fused-rings, and most of them are heptaketides (Supplementary Table S2). In this study, the phylogenetic analysis showed that groups II and III have a near homology relationship in evolutionary process, and the structural analysis indicated that the pocket volumes of PT domains of group III are bigger than those of group II. Similarly, group IV (C4-C9) has a closer homology relationship with group III than with group II, and the catalytic pockets of group IV possess larger cavity volumes than those of group III. Furthermore, the compounds synthesized by group IV have longer chains, such as the main metabolite, decaketide aflatoxin<sup>23,42,43</sup> (Supplementary Table S2).

In addition to the three common cyclizations, the rare cyclizations of C1-C6 and C3-C8 were also analyzed (Supplementary Figs. S2 & S3). Currently, there is only one case of C1-C6 aldol condensation in group V which catalyze the heptaketide cyclization. Structure modeling showed that the cavity volume of C1-C6 is merely about one half of the C6-C11 volume. The PT pocket of C1-C6 cyclization also lacks finger-like region and the catalytic dyad locates at the bottom of the pocket. So far, only two heptaketide synthases *P. aethiopicum* GsfA and *A. nidulans* PkgA in group V have been found to catalyze C1-C6 and C2-C7 heptaketide cyclizations, respectively. It is speculated that the cavity structures of group V cannot be suitable for the C6-C11 heptaketide cyclization, leading to the emergence of special C1-C6 and C2-C7 heptaketide cyclizations (Supplementary Table S2). The cavity shape of the sole case of C3-C8 cyclization is similar to that of C2-C7 in group I. However, due to its residue mutations, the cavity volume of C3-C8 is relatively small.

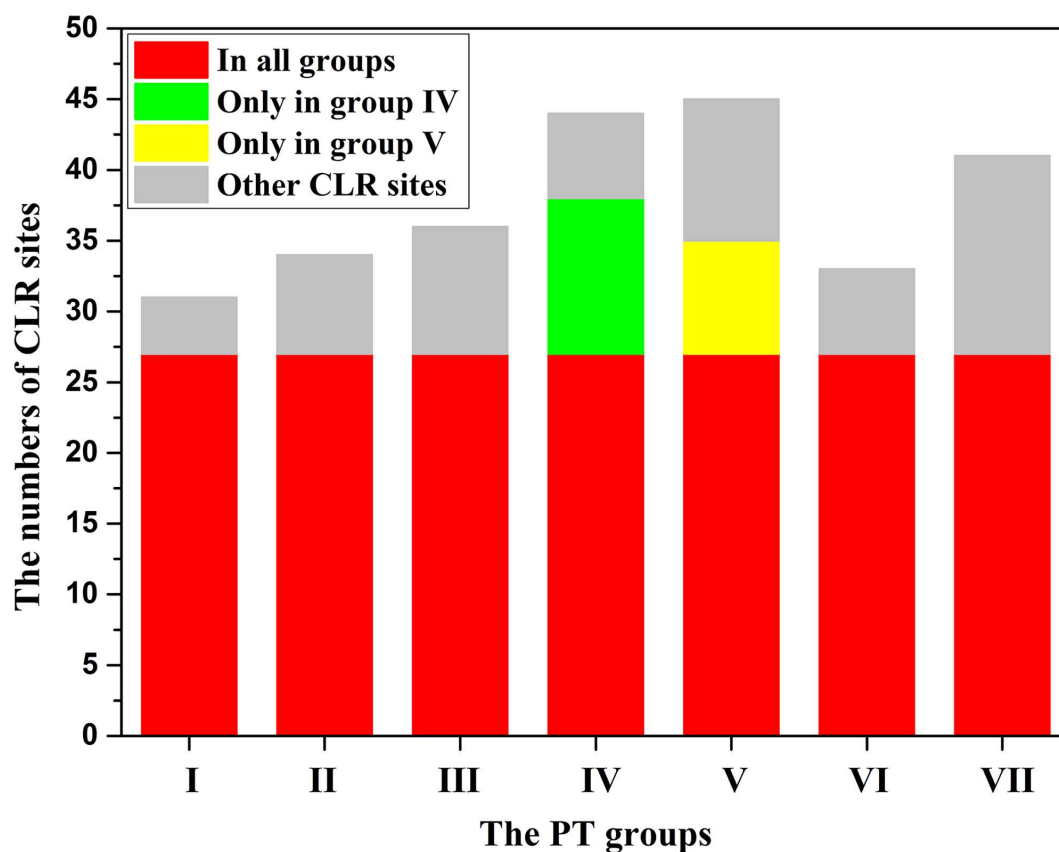
The TE/CLC domains in group V are absent reflected by the phylogenetic analysis of NR-PKs, while these NR-PKs still have the ability to generate multi-ring compounds (Supplementary Table S2). The previous research on crystal structure supported that the PT catalytic pocket of PksA in group IV can accommodate the regiospecific cyclizations for two rings<sup>12</sup>. According to the above analysis, the PT domains with the finger-like regions in group V have the larger cavity volumes, which may enable the PT domains to catalyze more cyclization processes.

**Sequence differentiation of PT domains.** The cavity lining residue (CLR) sites of each PT sequence were identified from the catalytic pockets selected by CASTp<sup>44</sup>, and then the CLR sites of each group were determined by multiple sequence alignment. The CLR sites of each group were considered only if they presented in the catalytic pockets at least twice in three proteins of each PT group. The CLR sites of seven groups (groups I-VII) were identified, except for group VIII with only one known regioselective cyclization sequence.

It was indicated that the catalytic pockets of each PT group are composed of 31-45 CLR sites (Table 1). The CLR site number and cavity volume of each PT group showed a positive correlation ( $r=0.93$ ,  $p<0.01$ ). Groups IV and V have the most CLR sites corresponding with the biggest cavity volumes. A total of 64 CLR sites appear in seven PT groups, including the CLR sites only in one group (Supplementary Table S5). Multiple sequence alignment revealed that among the predicted CLR sites, 27 CLR sites are common in seven groups (group I-VII), accounting for 87% of total CLR sites in group I (the most), and 60% of total CLR sites in group V (the minimum) (Fig. 5). These results suggested that most of the CLR sites are consistent in different groups, despite the various shapes and sizes of catalytic pockets.

The sequence differentiation of PT domains were also analyzed. The conservation analyses of all amino acid residues were performed with the evolutionary conservation scores calculated by ConSurf<sup>45</sup>. The evolutionary conservation of amino acid residues could be estimated with ConSurf based on the phylogenetic relations of homologous sequences. The analysis results showed that the proposed catalytic dyads (Asp 1543/His 1345 as in *A. parasiticus* PksA) are highly conserved.

Furthermore, most of the conserved residues in 661 PT sequences locate at the catalytic pockets (Fig. 6A, maroon residues). Based on the statistical analysis, the conservation of residues in the whole sequences is low (~35%, ConSurf grade 7-9), while for all of the CLR sites (a total of 64 sites, including the CLR sites appearing only in one group), the proportion of conserved residues reaches 83% (Fig. 6B). There are 25 residues (ConSurf grade 7-9) in the 27 common CLR sites are conserved, accounting for more than 92%. Particularly, there are 15 residues are the most conserved residues (ConSurf grade 9) in the common CLR sites. Therefore, the residues in CLR sites are relatively conserved in evolutionary process comparing with the residues in the whole PT sequences. These results suggested that the cyclization functions significantly constrain the evolution of PT sequences. In a previous study, the 6 residues in group IV, H1345, D1543, Q1547, N1554, T1546, and N1568 (number to *A. parasiticus* PksA prevail)



**Figure 5. Statistical analysis of CLR sites of PT groups (groups I–VII).** The red block, 27 common CLR sites in groups I–VII; the green block, 11 specific CLR sites in group IV; the yellow block, 8 specific CLR sites in group V; the gray block, the total CLR sites except for the common CLR sites and the specific CLR sites in groups IV and V.

belonging to the 27 common CLR sites have been validated to be important residues for activity by mutations<sup>12</sup>.

In groups IV and V, however, besides the 27 common CLR sites, 11 specific CLR sites and 8 specific CLR sites were found to only appear in the catalytic pockets of these two groups, respectively (Supplementary Table S5). These residue sites may associate with the finger-like regions of internal pockets of these two groups. The CLRs in group IV, G1491, M1495, M1498 and A1499 (number to *A. parasiticus* PksA prevail), are involved in the formation of finger-like region. In the corresponding positions of other groups, these CLRs are replaced by the bulkier residues (e.g., Val, Phe, and Leu), preventing the formation of finger-like region. The previous experimental data showed that the specific CLR G1491 was mutated to Leu, resulting in no apparent activity of the PT domain<sup>12</sup>. Additionally, in the catalytic pockets of group IV, the CLRs A1397 and L1622 replace the conserved Pro and Phe in other groups, respectively, resulting in the enlargement of pocket volumes. In group V, V1567 and G1638 (number to *A. nidulans* MdpG prevail) substitute the bulkier residues in other groups, including Phe and Ala, to generate finger-like region. Similarly, M1378 and G1423 in group V take the place of the conserved Val and Pro in other groups, respectively, resulting in the increase of the pocket volumes.

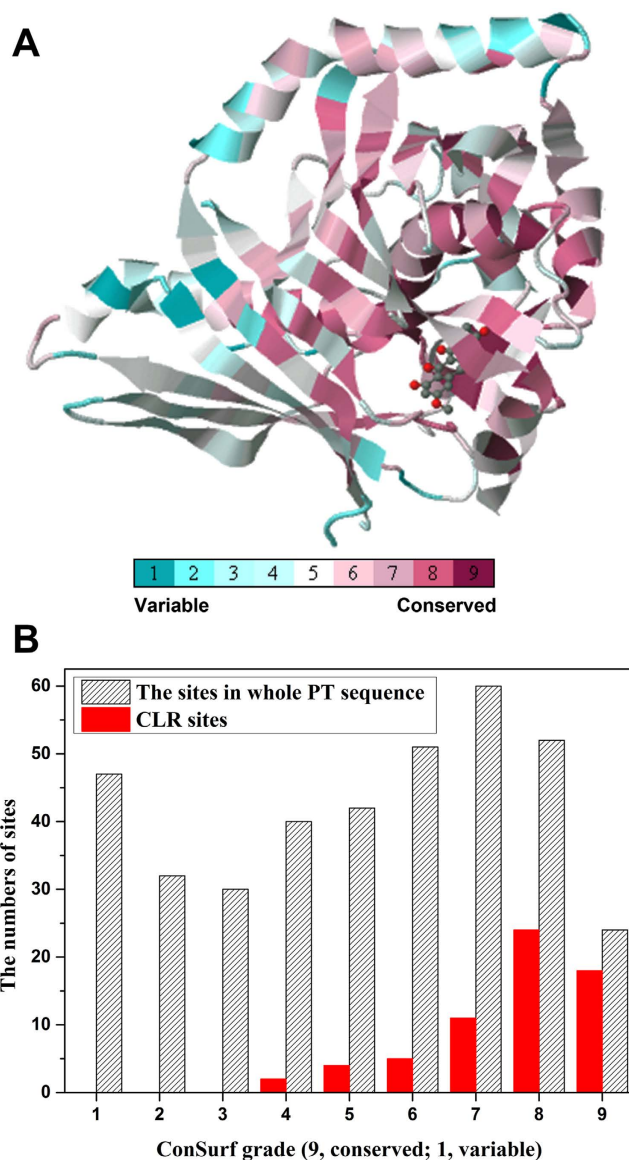
Furthermore, nearby the catalytic dyad, Thr (1354) in group IV and Thr (1385) in group V, are replaced by Pro in all C2–C7 groups. The spatial structure of Pro may influence the ring architecture and prevent the transferring of the first ring with C2–C7 regio-specific cyclization.

The above analysis suggested that the CLRs related with the catalysis are highly conserved in the main catalytic cavity of the pockets. The volume diversification of PT pockets affects the different cyclization regiospecificity and even compound formations. Residue mutation is the main reason to change the sizes and shapes of the PT pockets.

## Conclusions

In present study, the relationships between sequences, structures and functions of PT domains in fungal polyketide synthases were analyzed and elucidated. In the evolutionary process, the PT domains diverge into eight groups with functional distribution to synthesize diverse polyketides with variety of substrates





**Figure 6. Conservation scale analysis of residue sites in PT domains.** A. The structure of 3HRR HC8-bound structure belonging to Q12053. Conservation scale is defined from the most variable residue sites (grade 1, color represented by turquoise) which are considered as rapidly evolving to conservative residue sites (grade 9, color represented by maroon) which are considered as slowly evolving. B. The statistical analysis of PT sequence sites and CLR sites with ConSurf grades. The ConSurf grades were calculated with 661 PT sequences. The whole PT sequence was exemplified by Q12053 with 378 sites. The CLR sites were 64 in total from groups I–VII, including the CLR sites appearing only in one group.

by different cyclization modes. The functional differentiation is attributed to the changes of the cavity shapes and sizes with CLR mutations arising from the sequence variations of PT domains. The above findings may help to better understand the cyclization mechanisms of PT domains and even predict the structural types of the products especially with aromatic polyketide backbones from NR-PKS protein sequences. It could be prospected to manipulate the fungal PKS genes and further regulate the biosynthetic pathway to obtain the target metabolites.

## Methods

**Dataset.** The amino acid sequences of 55 fungal NR-PKSs with known cyclization modes were collected from NCBI database (Genbank)<sup>46</sup> with accession numbers which were obtained from the literature. The homologous sequences of 661 fungal NR-PKSs were searched and obtained by BLAST based on the 55 sequences<sup>47</sup>. The repetitive sequences and partial sequences were eliminated. The accession numbers and related information of NR-PKSs were provided in Supplementary Table S1 and Table S2.

| Group <sup>a</sup> | Regioselectivity cyclization | Accession number                   | Number of CLRs <sup>b</sup> |
|--------------------|------------------------------|------------------------------------|-----------------------------|
| I                  | C2-C7                        | XP_681178, ABB90282, ACD39762      | 31                          |
| II                 | C2-C7                        | BAA18956, AAN75188, ABU63483       | 34                          |
| III                | C2-C7                        | Q03149, AAU10633, EDP55264         | 36                          |
| IV                 | C4-C9                        | Q12053 (3HRR), Q12397, CCE67070    | 44                          |
| V                  | C6-C11                       | XP_657754, XP_663604, XP_001394705 | 45                          |
| VI                 | C2-C7                        | XP_681652, ADY00130, XP_664052     | 33                          |
| VII                | C2-C7                        | XP_658638, XP_660834, XP_001212610 | 41                          |

**Table 1.** The data of PT domain catalytic pockets. <sup>a</sup>Group VIII is not considered because only one sequence is studied with regioselective cyclization. <sup>b</sup>CLRs are present in the catalytic pockets at least twice in three proteins of each PT group.

The PT domain sequences of NR-PKSs were extracted and calibrated with UMA procedure<sup>48</sup>, SMART<sup>49</sup>, and CDD<sup>50</sup>.

**Phylogenetic Analysis.** The NR-PKS sequences and PT sequences were aligned with MAFFT<sup>51</sup>, respectively. Phylogenetic analyses were conducted using MEGA version 6 by the bootstrap neighbor joining method<sup>52</sup>. The evolutionary distances were computed using the Poisson correction method and were in the units of the number of amino acid substitutions per site. The phylogenetic tree was displayed by iTOL<sup>53</sup>.

**Structure modeling.** The three-dimensional models of PT domains were constructed using comparative protein modeling methods (Supplementary Table S3). The structural templates for PT domains were identified by BLAST search against PDB or by using threading approach. Three structural models for each PT domain were acquired by SWISS-MODEL<sup>35</sup>, I-TASSER<sup>36</sup> and PHYRE2<sup>37</sup>, respectively. All the structural models were refined in the atomic-level by the fragment-guided molecular dynamics (FG-MD) simulations<sup>38</sup>. Two forms of quality assessment, TM-score<sup>39</sup> and Ramachandran plot<sup>40</sup>, have been used to quantitatively assess the accuracy of protein structure predictions.

**Cavity volume and CLR site analysis.** The structural mapping and pocket architecture visualization were displayed using VMD<sup>54</sup>. The cavity volumes and the cavity lining residue (CLR) sites were calculated and identified by CASTp with 1.4 Angstroms probe radius<sup>44</sup>, respectively. The cavity volume of each PT protein was defined as the average value of the three models acquired by SWISS-MODEL+MD, I-TASSER+MD, and PHYRE2+MD. The CLR sites of each PT protein were defined as the sites that are present in the catalytic pockets at least twice in three models.

**Evolutional conservation analysis.** The evolutionary conservation of amino acid positions in the PT sequences was estimated by using ConSurf algorithm<sup>45</sup> (Supplementary Table S5). The LG substitution matrix and computation were based on the empirical Bayesian paradigm. Conservation scale was defined from the most variable amino acid positions (grade 1, color represented by turquoise) which were considered as rapidly evolving to conservative positions (grade 9, color represented by maroon) which were considered as slowly evolving.

## References

- Chooi, Y. H. & Tang, Y. Navigating the fungal polyketide chemical space: from genes to molecules. *J Org Chem.* **77**, 9933–9953 (2012).
- Crawford, J. M. & Townsend, C. A. New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol.* **8**, 879–889 (2010).
- Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl.* **48**, 4688–4716 (2009).
- Funa, N., Awakawa, T. & Horinouchi, S. Pentaketide resorcylic acid synthesis by type III polyketide synthase from *Neurospora crassa*. *J Biol Chem.* **282**, 14476–14481 (2007).
- Zhou, H., Li, Y. & Tang, Y. Cyclization of aromatic polyketides from bacteria and fungi. *Nat Prod Rep.* **27**, 839–868 (2010).
- Chiang, Y. M., Chang, S. L., Oakley, B. R. & Wang, C. C. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol.* **15**, 137–143 (2011).
- Crawford, J. M. *et al.* Deconstruction of iterative multidomain polyketide synthase function. *Science.* **320**, 243–246 (2008).
- Crawford, J. M., Dancy, B. C., Hill, E. A., Udvary, D. W. & Townsend, C. A. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc Natl Acad Sci USA.* **103**, 16728–16733 (2006).
- Li, Y. *et al.* Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J Biol Chem.* **285**, 22764–22773 (2010).
- 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.* **100**, 15670–15675 (2003).

11. Ahuja, M. *et al.* Illuminating the diversity of aromatic polyketide synthases in *Aspergillus nidulans*. *J Am Chem Soc.* **134**, 8212–8221 (2012).
12. Crawford, J. M. *et al.* Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature.* **461**, 1139–1143 (2009).
13. Vagstad, A. L., Hill, E. A., Labonte, J. W. & Townsend, C. A. Characterization of a fungal thioesterase having Claisen cyclase and deacetylase activities in melanin biosynthesis. *Chem Biol.* **19**, 1525–1534 (2012).
14. Reeves, C. D., Hu, Z., Reid, R. & Kealey, J. T. Genes for the biosynthesis of the fungal polyketides hypothemycin from *Hypomyces subiculosus* and radicicol from *Pochonia chlamyosporia*. *Appl Environ Microbiol.* **74**, 5121–5129 (2008).
15. Wang, S. *et al.* Functional characterization of the biosynthesis of radicicol, an Hsp90 inhibitor resorcylic acid lactone from *Chaetomium chiversii*. *Chem Biol.* **15**, 1328–1338 (2008).
16. Gaffoor, I. & Trail, F. Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Appl Environ Microbiol.* **72**, 1793–1799 (2006).
17. Schroeckh, V. *et al.* Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci USA.* **106**, 14558–14563 (2009).
18. Xu, Y. *et al.* Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat shock response-modulating anticancer fungal polyketide from *Aspergillus terreus*. *Appl Environ Microbiol.* **79**, 2038–2047 (2013).
19. Fujii, I. *et al.* Enzymatic synthesis of 1,3,6,8-tetrahydroxynaphthalene solely from malonyl coenzyme A by a fungal iterative type I polyketide synthase PKS1. *Biochemistry.* **39**, 8853–8858 (2000).
20. Watanabe, A. *et al.* *Aspergillus fumigatus alb1* encodes naphthopyrone synthase when expressed in *Aspergillus oryzae*. *FEMS Microbiol Lett.* **192**, 39–44 (2000).
21. Wiemann, P. *et al.* Biosynthesis of the red pigment bikaverin in *Fusarium fujikuroi*: genes, their function and regulation. *Mol Microbiol.* **72**, 931–946 (2009).
22. Fujii, I. *et al.* Hydrolytic polyketide shortening by *ayg1p*, a novel enzyme involved in fungal melanin biosynthesis. *J Biol Chem.* **279**, 44613–44620 (2004).
23. Feng, G. H. & Leonard, T. J. Characterization of the polyketide synthase gene (*pksL1*) required for aflatoxin biosynthesis in *Aspergillus parasiticus*. *J Bacteriol.* **177**, 6246–6254 (1995).
24. Yu, J. H. & Leonard, T. J. Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase. *J Bacteriol.* **177**, 4792–4800 (1995).
25. Choquer, M. *et al.* The *CTB1* gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of *Cercospora nicotianae*. *Mol Plant Microbe Interact.* **18**, 468–476 (2005).
26. Awakawa, T., Kaji, T., Wakimoto, T. & Abe, I. A heptaketide naphthaldehyde produced by a polyketide synthase from *Nectria haematococca*. *Bioorg Med Chem Lett.* **22**, 4338–4340 (2012).
27. Chooi, Y. H., Cacho, R. & Tang, Y. Identification of the viridicatumtoxin and griseofulvin gene clusters from *Penicillium aethiopicum*. *Chem Biol.* **17**, 483–494 (2010).
28. Boruta, T. & Bizukojc, M. Culture-based and sequence-based insights into biosynthesis of secondary metabolites by *Aspergillus terreus* ATCC 20542. *J Biotechnol.* **175**, 53–62 (2014).
29. Szewczyk, E. *et al.* Identification and characterization of the asperthecin gene cluster of *Aspergillus nidulans*. *Appl Environ Microbiol.* **74**, 7607–7612 (2008).
30. Li, Y., Chooi, Y. H., Sheng, Y., Valentine, J. S. & Tang, Y. Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an alpha-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase. *J Am Chem Soc.* **133**, 15773–15785 (2011).
31. Regueira, T. B. *et al.* Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Appl Environ Microbiol.* **77**, 3035–3043 (2011).
32. Shimizu, T. *et al.* Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl Environ Microbiol.* **71**, 3453–3457 (2005).
33. Balakrishnan, B. *et al.* Genetic localization and *in vivo* characterization of a *Monascus azaphilone* pigment biosynthetic gene cluster. *Appl Microbiol Biotechnol.* **97**, 6337–6345 (2013).
34. Lackner, G., Bohnert, M., Wick, J. & Hoffmeister, D. Assembly of melleolide antibiotics involves a polyketide synthase with cross-coupling activity. *Chem Biol.* **20**, 1101–1106 (2013).
35. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–258 (2014).
36. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* **12**, 7–8 (2014).
37. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* **4**, 363–371 (2009).
38. Zhang, J., Liang, Y. & Zhang, Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure.* **19**, 1784–1795 (2011).
39. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography.* **26**, 283–291 (1993).
41. Chiang, Y. M. *et al.* A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *J Am Chem Soc.* **131**, 2965–2970 (2009).
42. Tominaga, M. *et al.* Molecular analysis of an inactive aflatoxin biosynthesis gene cluster in *Aspergillus oryzae* RIB strains. *Appl Environ Microbiol.* **72**, 484–490 (2006).
43. Cary, J. W., Ehrlich, K. C., Beltz, S. B., Harris-Coward, P. & Klich, M. A. Characterization of the *Aspergillus ochraceoroseus* aflatoxin/sterigmatocystin biosynthetic gene cluster. *Mycologia.* **101**, 352–362 (2009).
44. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–118 (2006).
45. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–533 (2010).
46. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
47. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–9 (2008).
48. Udwy, D. W., Merski, M. & Townsend, C. A. A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J Mol Biol.* **323**, 585–598 (2002).
49. Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–305 (2012).
50. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–352 (2013).

51. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**, 772–780 (2013).
52. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* **30**, 2725–2729 (2013).
53. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–478 (2011).
54. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph.* **14**, 33–38, 27–38 (1996).

### Acknowledgments

We thank Prof Townsend C.A. from the Department of Chemistry, Johns Hopkins University for providing the Udwy–Merski algorithm. We appreciate Dr. Han W.J. and Dr. Han K. from the State Key Laboratory of Microbial Technology, Shandong University for technical advice and beneficial discussions. This work was supported by the National Natural Science Foundation of China (Nos. 41130858; 41322037; 41376145) and the Fundamental Research Funds for the Central Universities of China (No. 201122005).

### Author Contributions

L.L. and Z.Z. designed the experiments, analyzed the data, and wrote the paper; J.L.W. and H.B. collected the data and revised the manuscript; C.L.S and C.Y.W directed the experiments, and wrote and revised the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Liu, L. *et al.* Bioinformatical Analysis of the Sequences, Structures and Functions of Fungal Polyketide Synthase Product Template Domains. *Sci. Rep.* **5**, 10463; doi: 10.1038/srep10463 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>