

## 研究论文

DOI: 10.12211/2096-8280.2021-005

## 虫生真菌非核糖体多肽活性产物生物合成潜力预测

张礼文<sup>1</sup>, István MOLNÁR<sup>2</sup>, 徐玉泉<sup>1</sup>

(<sup>1</sup> 中国农业科学院生物技术研究所, 北京 100081; <sup>2</sup> 美国亚利桑那大学西南天然产物研究中心, 亚利桑那州 图森 85706, 美国)

**摘要:** 肉座菌目虫生真菌合成的非核糖体多肽类天然产物具有抗菌、杀虫、抗癌、调节免疫等生物活性, 在临床和农业等领域有重要应用价值。近年来, 随着真菌基因组测序数量的迅速增加、注释和分析工具的不断进步, 人们发现真菌基因组中存在大量产物未知的天然产物合成基因。准确有效地预测这些基因的功能, 筛选最有潜力合成新颖天然产物的基因簇, 可以提高天然产物挖掘的效率。本研究选取40株肉座菌目虫生真菌基因组, 系统分析了编码非核糖体多肽合成酶的基因及基因簇。基于隐马尔可夫模型, 预测了445个模块型非核糖体多肽合成酶和1243个类非核糖体多肽合成酶; 通过提取腺苷酰化结构域, 构建序列相似性网络, 以已知功能的非核糖体多肽合成酶作为标签, 利用马尔可夫聚类算法, 分析了非核糖体多肽产物的主要类别, 发现了可能合成线性多肽、环缩肽、脂多肽、生物碱等新型结构化合物的基因簇。研究结果不仅揭示了肉座菌目虫生真菌合成非核糖体多肽活性产物的巨大潜力, 而且为通过激活沉默基因簇挖掘新产物、利用合成生物学手段改造合成途径提供参考。

**关键词:** 肉座菌目虫生真菌; 非核糖体多肽合成酶; 非核糖体多肽; 生物合成基因簇; 序列相似性网络分析

**中图分类号:** Q74 **文献标志码:** A

## Potential biosynthesis of nonribosomal peptides by hypocrealean entomopathogenic fungi

ZHANG Liwen<sup>1</sup>, István MOLNÁR<sup>2</sup>, XU Yuquan<sup>1</sup>

(<sup>1</sup>Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; <sup>2</sup>Southwest Center for Natural Products Research, University of Arizona, Tucson, AZ 85706, USA)

**Abstract:** Nonribosomal peptide natural products synthesized by Hypocrealean entomopathogenic fungi have antibacterial, insecticidal, anticancer, immunomodulatory and other biological activities, with high potential for application in clinical or agricultural fields. These bioactive compounds are synthesized by nonribosomal peptide synthetases (NRPSs) and tailored by additional enzymes that are encoded by clustered genes. In addition to the 20 amino acids, non-canonical amino acids as well as  $\alpha$ -hydroxy acids can also be incorporated into nonribosomal

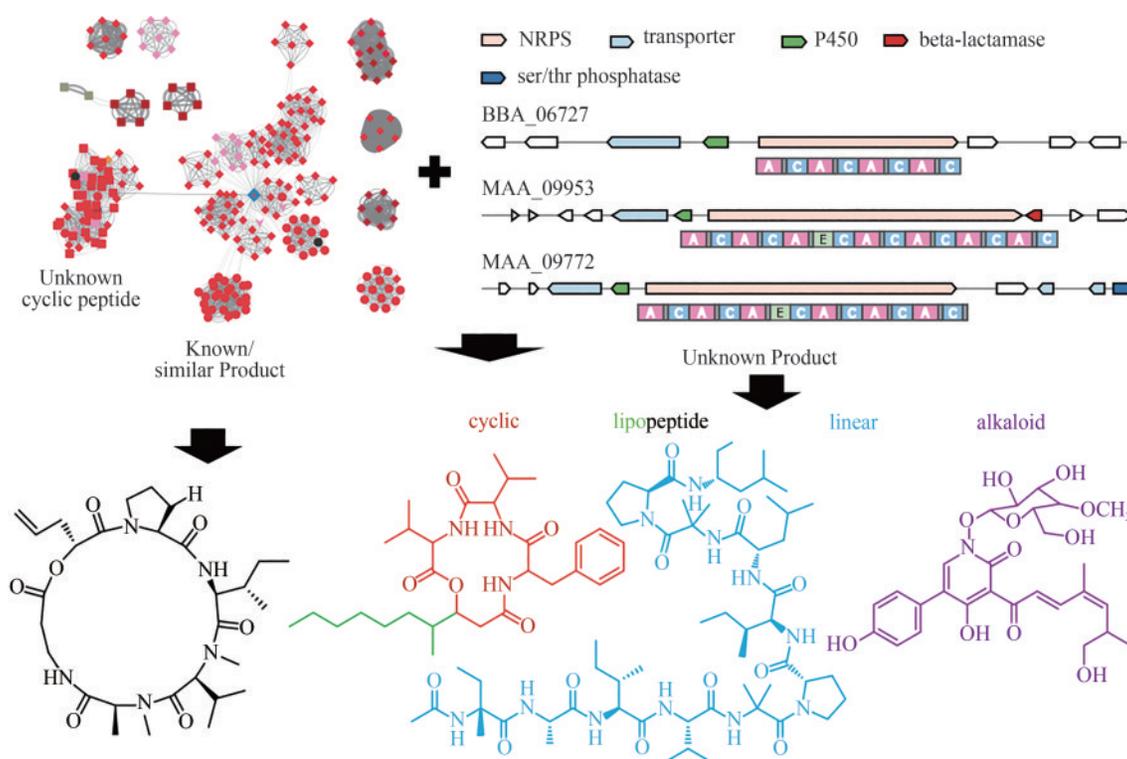
收稿日期: 2021-01-12 修回日期: 2021-04-16

基金项目: 国家自然科学基金 (32070064, 31870076, 32070053); 中国农业科学院基本科研业务费专项 (Y2020XK20); 中国农业科学院科技创新工程

引用本文: 张礼文, István MOLNÁR, 徐玉泉. 虫生真菌非核糖体多肽活性产物生物合成潜力预测[J]. 合成生物学, 2021, 2(5): 815-825

Citation: ZHANG Liwen, István MOLNÁR, XU Yuquan. Potential biosynthesis of nonribosomal peptides by hypocrealean entomopathogenic fungi[J]. Synthetic Biology Journal, 2021, 2(5): 815-825

peptides as structural units, and these, together with further modifications, empower an almost unlimited structural diversity. The rapid increase in the number of sequenced fungal genomes shows that there exists a large number of nonribosomal peptide biosynthetic gene clusters with unknown functions. Accurate and effective prediction of the functions of these "orphan" biosynthetic gene clusters can help to select those clusters that have potential to synthesize novel natural products, and increase the efficiency of natural product genome mining. In this study, we systematically analyzed the genes encoding NRPSs and their gene clusters in the genomes of 40 strains from Hypocrealean entomopathogenic fungi. Genes encoding NRPSs were predicted and categorized based on the hidden Markov models for adenylation, condensation and thiolation domains to reveal 445 modular NRPSs and 1243 NRPS-like proteins. A sequence similarity network based on the amino acid sequences of adenylation domains of these synthetases was constructed. Using adenylation domains from known NRPSs as references, we analyzed the main categories of predicted nonribosomal peptide products using the Markov clustering algorithm. We identified several biosynthetic gene clusters that potentially yield known bioactive compounds or their congeners. In addition, biosynthetic gene clusters that may be able to produce new bioactive compounds belonging to the linear peptide, cyclic peptide, lipopeptide, and alkaloid structural classes were also discovered. The results reveal a great potential of Hypocrealean entomopathogenic fungi to synthesize nonribosomal peptides, and provide insight for genome mining to identify new products by activating silent gene clusters or modifying biosynthetic pathways by synthetic biology methods.



**Keywords:** hypocrealean entomopathogenic fungi; nonribosomal peptide synthetase (NRPS); nonribosomal peptide (NRP); biosynthetic gene clusters (BGCs); sequence similarity network analysis (SSN)

通过测序和注释,发现真菌基因组中存在大量参与非核糖体多肽天然产物生物合成的基因

簇<sup>[1]</sup>,具有合成结构多样化化合物的巨大潜力<sup>[2]</sup>。大多数真菌非核糖体多肽合成酶(nonribosomal

polypeptide synthetase, NRPS) 包括多个由腺苷酰化 (adenylation, A)、巯基化 (thiolation, T) 和缩合 (condensation, C) 等结构域组成的模块。在生物合成的过程中, 腺苷酰化结构域先将氨基酸或羟酸前体单元加载到巯基化结构域上, 然后缩合结构域缩合前体单元和中间产物合成非核糖体多肽<sup>[3]</sup>。除了这3个核心结构域, 非核糖体多肽合成酶还可能包含修饰前体单元或中间产物甲基化、差向异构化的结构域, 以及终止延伸和释放产物、末端缩合等功能的结构域, 从而使合成的产物结构更加复杂多样。

肉座菌目虫生真菌能够产生大量具有抗菌、杀虫、抗癌、调节免疫等活性的天然产物, 如环孢霉素和白僵菌素等, 具有重要药用和农用价值<sup>[3-8]</sup>。通过基因缺失和互补、异源表达、化合物结构解析等方法, 阐明了部分肉座菌目虫生真菌天然产物生物合成基因簇的功能, 如合成破坏素 (destruxins)、白僵菌素 (beauvericin)、白僵菌环四肽 (bassianolide) 和 serinocyclins 等<sup>[9-13]</sup>, 然而, 很多基因簇在常规实验室条件下不表达, 因此绝大多数基因簇的产物未知<sup>[14-15]</sup>。在过去数年中, 使用特定方法激活沉默基因簇的表达在发现新型真菌天然产物方面虽然取得了重大进展<sup>[1, 16-17]</sup>, 然而, 面对数以千万计的生物合成基因簇, 通过实验的手段鉴定其功能十分费时费力。如何高效选取最具合成新颖天然产物潜力的基因簇成为亟待解决的问题。因此, 本研究基于隐马尔可夫模型从虫生真菌基因组中预测可能合成非核糖体多肽类天然产物的非核糖体多肽合成酶, 通过构建序列相似性网络将其分类, 并与已知产物的标签结构域进行比较, 从而推测出可能合成新颖产物的生物合成基因簇。准确有效预测基因乃至基因簇的功能, 不仅可以指导大规模实验鉴定和筛选, 而且也后续利用组合生物合成改造天然途径提供了候选的基因元件。

## 1 材料与方法

### 1.1 数据集和标签结构域的收集

本研究使用来自 SwissProt (2019年发行版) 审核和人工注释的全部真菌蛋白作为标签蛋白。40种肉座菌目虫生真菌的基因组序列和预测的蛋白序列来自于 NCBI 数据库。使用基于隐马尔可夫模型 (hidden Markov model, HMM) 的 HMMSearch 方法预测腺苷酰化 (PFAM00501.21) 和缩合 (PFAM00668.13) 等结构域 (表 1), 并使用 Python 提取相应的序列。来自于 SwissProt 已知功能的结构域被定义为“标签”结构域。

### 1.2 序列相似网络的建立

为了建立序列相似性网络, 使用 BLASTp 程序计算数据集中每一对腺苷酰化结构域的氨基酸序列一致性 (identity)、覆盖率 (coverage) 和 e-value。经统计分析得出, 一致性数据在 20%~60% 区间符合正态分布的规律, 高于 60% 则逐渐偏离正态分布, 离群值的下边界为 62.06%, 这表明当两个腺苷酰化结构域的氨基酸序列一致性高于 62.06% 时为“真”相似的概率较高。因此在构建网络和聚类时, 将一致性 0 参数的阈值设定为 62.06%。类似地, 覆盖率值通常在 91%~99% 之间, 低于 75% 为异常值, 因此覆盖率的阈值设定为 75%。图形可视化使用 Cytoscape ver 3.7.2。利用马尔可夫聚类算法 (MCL), 根据序列相似性矩阵来识别相关节点的分支。再根据 NRPS 的结构域组成, 并结合 MultiGeneBlast 评估的特定分支中基因簇的相似性, 来进一步分析特定分支的特点并预测可能的代谢产物。

表1 非核糖体多肽合成酶结构域注释

Tab. 1 Annotion for NRPS domains in hypocrealean entomopathogenic fungi (HEF)

Description	Function with HMM	Source	Cutoff (Score)	Length	Domains (SwissProt)	Domains (HEF)
Adenylation domain	AMP-binding	PFAM00501.21	55	418	280	2297
Condensation domain	Condensation	PFAM00668.13	50	301	255	1483
Thiolation domain	PP-binding	PFAM0050.20	30	67	312	2775

## 2 肉座菌目虫生真菌基因组测序及次生代谢产物合成基因簇概况

2014—2019年期间,已发表的肉座菌目虫生真菌基因组序列数量大幅增加:2014年以前为8个,2014年以后增至31个,这为分析天然产物的合成潜力、比较基因簇之间的异同提供了丰富的数据。本研究选取的40种真菌来自于白僵菌属(*Beauveria*)、冬虫夏草属(*Cordyceps*和*Ophiocordyceps*)、被毛孢属(*Hirsutella*)、座壳孢属(*Hypocrella*或*Aschersonia*)、棒束孢属(*Isaria*)、轮枝孢属(*Lecanicillium*)、莫勒菌属(*Moelleriella*)、弯颈霉属(*Tolyocladium*)等,为农业上应用广泛的生防真菌,也是常用的药用真菌(菌种目录详见<http://www.researchgate.net/publication/354887574>\_Table S1)。

通过HMMsearch检索所得的腺苷酰化结构域、缩合结构域和非核糖体多肽合成酶数量见表1、表2。腺苷酰化结构域的氨基酸序列一致性的25%分位数为27.70%,75%分位数为41.30%,离群值的下边界为62.06%。覆盖率值通常在91%~99%之间,异常值低于75%。因此,在后续分析中,分别选择62.06%和75%作为一致性和覆盖率的阈值,从而得到了182个腺苷酰化结构域的分支(clade)。

表2 非核糖体多肽合成酶统计信息

Tab. 2 Statistic summary of NRPSs in hypocrealean entomopathogenic fungi (HEF)

Subtype	Definition (Proteins that contain...)	Number (SwissProt)	Number (HEF)
Multi-modular	more than one A AND C domains	57	280
Single-modular	ONE A AND C domains	21	165
Pseudo	A OR C domain	86	1243

注:统计结果不含有PKS-NRPS杂合酶

Note: The PKS-NRPS hybrid enzymes were excluded.

## 3 非核糖体多肽合成酶序列相似性网络和马尔可夫聚类分析

序列相似性网络(sequence similarity network, SSN)能够通过同源蛋白之间成对的序列相似关

系,自动化评估蛋白的分类,并从家族/亚家族之间的关系来研究特定蛋白的功能。在SSN中,每个蛋白用一个符号(“节点”)表示;如果两个节点的序列相似性超过指定阈值,则通过一条线(“边”)连接。已有多篇文章综述了SSN在蛋白质序列与功能研究中的成功应用<sup>[18-20]</sup>。

腺苷酰化和缩合结构域均已成功用于非核糖体多肽合成酶系统发育研究<sup>[4, 21-26]</sup>。本研究在网络构建和聚类分析腺苷酰化结构域的基础上,分析了部分非核糖体多肽合成酶的模块和结构域组成,以及所在的基因簇特征。研究结果显示肉座菌目虫生真菌的基因组中广泛存在功能未知的非核糖体多肽合成酶家族,即使在亲缘关系十分接近的种属之间,它们在基因或基因簇水平上也存在高度变异性,表明肉座菌目虫生真菌是合成新型非核糖体多肽类天然产物的潜在资源。

大多数非核糖体多肽合成酶具有多个腺苷酰化结构域,因此,利用对腺苷酰化结构域进行聚类分析来阐明合成酶的功能并预测产物结构变得十分复杂。在真菌聚酮合酶中,酮脂酰合酶(ketoacyl synthase, KS)结构域往往可以代表整个酶的特征;与此相对的,非核糖体多肽合成酶被视为具有不同系统发育特征的模块复合体,而腺苷酰化结构域的聚类则代表其所在模块的特征。收集到的2676个腺苷酰化结构域(包括302个已知功能的标签结构域)的氨基酸序列平均一致性为36.30%,其中2358个被成功地分为174个分支(不含聚酮合酶-非核糖体多肽合成酶杂合酶中的腺苷酰化结构域)。由于数目较多,本文图中仅列出具有5个以上节点的分支簇,且排除了参与短链脂肪酸合成的酰基辅酶A连接酶(<http://www.researchgate.net/publication/354887574>\_Table S1)。由此产生的网络可将腺苷酰化结构域分为多模块、双模块、单模块和杂合的非核糖体多肽合成酶。这与先前报道的系统发育分析基本一致,即真菌非核糖体多肽合成酶分为两大分支:远古起源的单/双模块和新近起源的多/双模块<sup>[24, 27-28]</sup>。

### 3.1 具有新型非核糖体多肽合成潜力的多模块非核糖体多肽合成酶

肉座菌目虫生真菌基因组预测的2440个腺苷

酰化结构域中，有 1056 个来自于多模块的非核糖体多肽合成酶（图 1）<sup>[9, 11-14, 29-30]</sup>，这类合成酶仅发现于真菌基因组，其腺苷酰化结构域可能由垂直遗传、基因重排（如复制、插入和丢失）或水平转移事件进化而来<sup>[4, 24]</sup>，是发现新型活性产物的宝贵资源。

这些多模块非核糖体多肽合成酶中，绝大多

数的产物是未知的。如图 1 所示，第 52~64 分支的非核糖体多肽合成酶形成了一个较大的超级分支（super clade）。根据氨基酸序列相似性，这些合成酶的最后一个结构域属于环肽类的终端缩合结构域，可以通过分子内环化释放延伸的肽链，因此推测其产物是环肽类化合物。其中，52~60 分支可以进一步划分为 3 种含有 4 个以上模块的非核糖体多

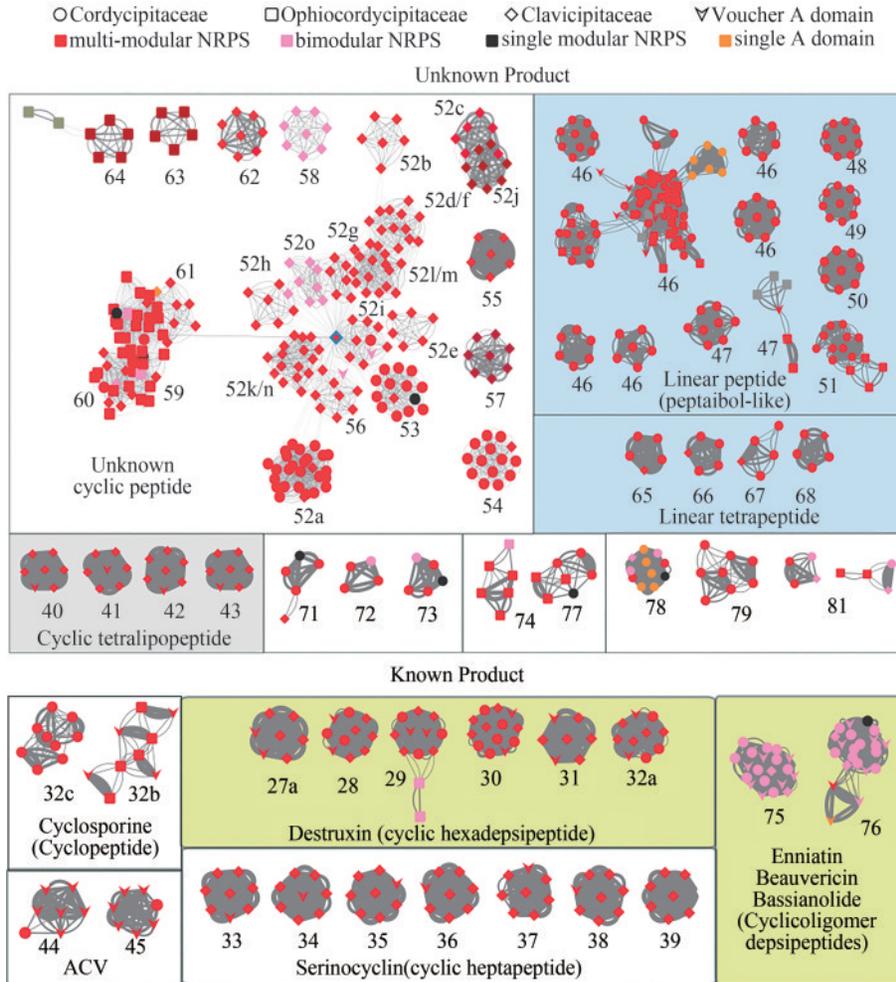


图 1 肉座菌目虫生真菌多模块非核糖体多肽合成酶的相似性网络图

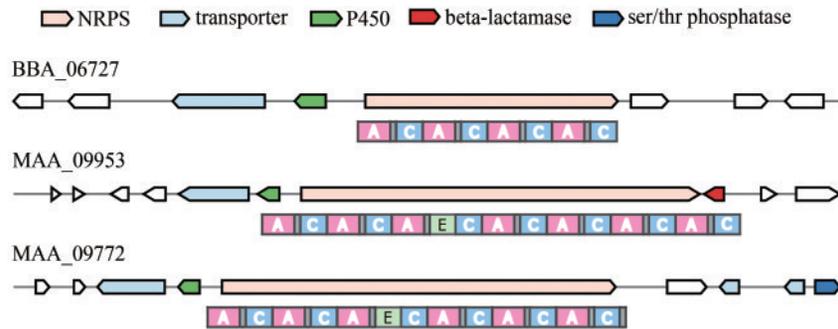
[其中每个节点代表一个腺苷酰化结构域，连接线粗细与相似性成正比。节点颜色显示非核糖体多肽合成酶的类型，形状表示肉座菌目虫生真菌的种类（科），已知功能的结构域用箭头表示。灰色边框中显示来自同一酶的腺苷酰化结构域的分支（如 DtxS1），或由密切相关的结构域分支（如 52~64）形成的超级分支（详见正文）。底色显示预测的产物类型。腺苷酰化结构域的亚分支用数字字母组合表示，例如 52a]

Fig. 1 Network for the subgroups of multi-modular NRPSs in Hypocrealean Entomopathogenic fungi

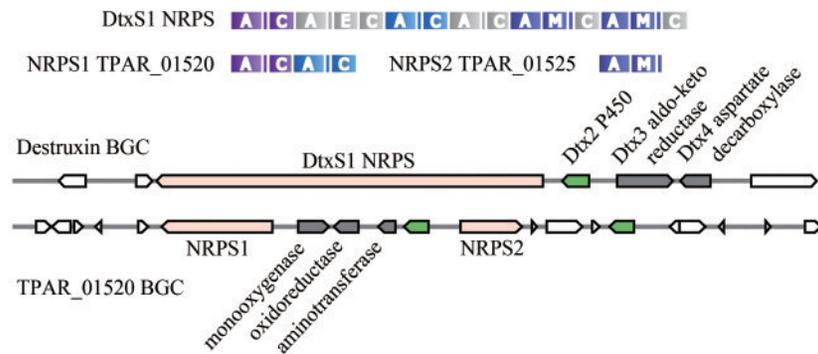
(Nodes in the network represent A domains, and the line boldness is proportional to the identity. The nodes are colored according to the domain composition of NRPS. Referenced A domains from voucher NRPSs with known products are indicated by arrowheads. The predicted products/functions of the enzymes are listed below the clades. Boxed groups show A domains from the group of enzymes that also contain DtxS1, or superclades that are formed by closely related A domain clades such as clades #52~64 (see text for details). Sub-clades of A domains are indicated with a number-letter combination, for example 52a. NRPSs of the BGCs that contain the voucher A domains include: #27a~32a, destruxin synthetase DtxS1<sup>[9]</sup>; #32b, cyclosporine synthetase<sup>[29]</sup>; #33~39, serinocyclin synthetase<sup>[13-14]</sup>; #44~45, ACVS: *N*-(5-amino-5-carboxypentanoyl)-L-cysteiny-D-valine synthetase in penicillin/cephalosporin biosynthesis; #46~51, Tex1: a peptaibol synthetase from *Trichoderma virens*<sup>[30]</sup>; #75~76, beauvericin and bassianolide synthetase<sup>[11-12]</sup>.)

肽合成酶，其代表为 *B. bassiana* 的 BBA\_06727、*M. robertsii* 的 MAA\_09953 和 MAA\_09772 (图 1)。通过系统发育分析，这些非核糖体多肽合成酶曾被归为“昆虫病原菌特有”的进化分支<sup>[3-4]</sup>。与 BBA\_06727 相似的蛋白有 13 个，结构域组成均为 (AT)-(CAT)-(CAT)-(CAT)-(C<sub>T</sub>)，它们的第 1 个腺苷酰化结构域归于第 53 个分支，第 3 个腺苷酰化结构域归于第 54 分支，第 2 个和第 4 个腺苷酰化结构域均归于分支 52a。同一个非核糖体多肽合成酶中的两个腺苷酰化结构域聚类在一起通常表明它们来源于进化上的复制事件。这一类非核糖体多肽合成酶所在的基因簇中还含有编码细胞色素 P450 单加氧酶和转运蛋白的基因 [图 2(a)]。BBA\_06727 的转录在昆虫感染期间被激活<sup>[27]</sup>，表明可能在昆虫致病过程中起作用。52b~52h 亚分支和第 55 分支的非核糖体多肽合成酶 (如 MAA\_09953) 存在于 6 种绿僵菌属真菌基因组中，其结构域组成均为 (AT)-(CAT)-(CATE)-(CAT)-

(CAT)-(CAT)-(CAT)-(C<sub>T</sub>)，所在基因簇中还包含编码细胞色素 P450 单加氧酶、β-内酰胺酶和转运蛋白的基因。以 MAA\_09772 为代表的非核糖体多肽合成酶在 8 种绿僵菌属真菌中十分保守，具有 (AT)-(CAT)-(CATE)-(CAT)-(CAT)-(CAT)-(C<sub>T</sub>T) 的结构域组成，其腺苷酰化结构域归于分支 59、亚分支 52i~52n、分支 60，相应的合成基因簇包含细胞色素 P450 单加氧酶、Ser/Thr 蛋白磷酸酶和两个转运蛋白的基因。亚分支 52o 和分支 58 的腺苷酰化结构域也来源于相同的非核糖体多肽合成酶，其结构域的组成为 (AT)-(CATE)-(CT)-(C<sub>T</sub>)，基因簇含有编码细胞色素 P450 单加氧酶、两种氨基转移酶、脱羧酶、双加氧酶和含有腺苷酰化结构域的连接酶/合成酶的基因。值得注意的是，超级分支 52 中几乎所有的分支都是通过 52i 中的一个节点连接的，这个节点对应于 *M. rileyi* 基因组中 NOR\_07696 的第 2 个腺苷酰化结构域 (介数 betweenness 为 0.826)，表明这个腺苷酰化结构域与共



(a) 超分支 52 中 3 个未知产物的多模块非核糖体多肽合成酶基因簇  
(a) Representative biosynthetic gene clusters containing NRPSs from superclade 52



(b) 比较破坏素合成酶 DtxS1 和 TPAR\_01520(NRPS1)、TPAR\_01525(NRPS2) 的结构域与基因簇  
(b) Domain and BGC comparisons between the destruxin synthetase DtxS1 and TPAR\_01520(NRPS1) and TPAR\_01525(NRPS2) of *T. paradoxum*

图 2 肉座菌目虫生真菌基因簇分析

Fig. 2 Analysis of representative biosynthetic gene clusters in hypocrealean entomopathogenic fungi

同的祖先最为相似。分支 59~61 中来自于线虫草科 (Ophiocordycipitaceae) 和麦角菌科 (Clavicipitaceae) 的非核糖体多肽合成酶也通过 NOR\_07696 相连。

分支 46~51 对应的非核糖体多肽合成酶可能产生 peptaibols 类线性多肽<sup>[30]</sup>, 其前体单元除氨基酸外, 还包括由基因簇中其他酶合成的非常规前体。这些多模块非核糖体多肽合成酶大多来自 *Trichothecium* 和 *Tolyposcladium* 属真菌<sup>[30]</sup>, 由 8 个以上 CAT 模块组成, 有的还具有特殊模块或结构域, 如 *T. inflatum* 的 ctg24\_orf003 具有一个酮酯酰合成酶和丙二酰/酰基转移酶结构域组成的加载模块。*O. australis* 的 CDD81\_5223 的终端结构域为硫酯酶结构域, 而不是 C<sub>T</sub>。这些非核糖体多肽合成酶所在的基因簇并不保守, 普遍存在转氨酶基因, 可能用于合成非常规氨基酸前体。

当非核糖体多肽合成酶两两比较时, 经常发现它们的部分腺苷酰化结构域能够聚类在一起, 而其余腺苷酰化结构域则属于不同的分支。例如, 来自 *T. paradoxum* 的 TPAR\_05120 对应于绿僵菌属破坏素合成酶 DtxS1 的第 1 和第 3 个结构域, 而 TPAR\_01525 则与 DtxS1 的第 5 或第 6 模块相似 [图 2(b)]。这种嵌套关系展示了非核糖体多肽合成酶复杂的进化过程, 使其产物有不同的结构和生物学功能。

除此之外, 一些分支包含产物已知的非核糖体多肽合成酶, 具有相同的结构域组成, 所在的基因簇也高度相似, 可能产生与已知化合物结构类似的产物。其中, 分支 75~76 包含白僵菌素/白僵菌环四肽合成酶的腺苷酰化结构域<sup>[11, 31]</sup>, 分支 27~32 和 33~39 则分别包含破坏素<sup>[9-10]</sup> 和 serinocyclin<sup>[13]</sup> 合成酶的腺苷酰化结构域 (图 1), 说明这些分支中的菌株具有合成杀虫化合物的潜力。

### 3.2 较为保守的单模块和多模块非核糖体多肽合成酶

如图 3 所示, 单/双模块的非核糖体多肽合成酶包括赖氨酸生物合成途径中还原  $\alpha$ -氨基己二酸的 L-氨基己二酸半醛脱氢酶 (分支 23)<sup>[24]</sup>, 参与形态发育但产物未知的 ChNPS10 合成酶 (分支 15)<sup>[40]</sup>, 产生聚硫二酮哌嗪的 ChNPS11/ETP 毒素

合成酶 (分支 12 和 13) 等<sup>[41]</sup>, 其合成的代谢物参与了真菌重要的生物过程, 如性发育、生殖、分生孢子发育等, 因此在基因组中也较为保守<sup>[24, 32-39]</sup>。其中, 部分基因的生理功能已有研究, 但其合成产物仍然未知。

铁载体是一种低分子量的铁螯合物, 可分为胞外和胞内两种类型, 因与病原真菌的致病性相关而受到广泛关注<sup>[37]</sup>。铁载体合成酶具有保守的结构域, 合成产物的结构和生物学功能相似, 在各种营养模式的真菌中广泛分布, 包括动植物病原菌和腐生菌。铁载体合成酶是多模块的非核糖体多肽合成酶, 在系统发育中属于较新的分支, 然而由于其序列和功能的保守性, 将它们与单/双模块的非核糖体多肽合成酶一起分析。大部分肉座菌目虫生真菌的基因组中都有铁载体合成酶 NPS2/SidC (分支 1~4)<sup>[36]</sup>、NPS6 (分支 9)<sup>[39]</sup>、SidN (分支 7 和 8)<sup>[38]</sup> 和/或 SidI (分支 21)<sup>[37]</sup> 的编码基因。值得注意的是, 分支 5 和 6 的腺苷酰化结构域与 NPS2/SidC 的第 1 个腺苷酰化结构域具有很近的亲缘关系, 但其所属的非核糖体多肽合成酶与 NPS2/SidC 截然不同: 仅有一个模块, 几乎只存在于白僵菌/冬虫夏草属真菌中, 与醇脱氢酶、细胞色素 P450 单加氧酶、聚酮合酶或聚酮合酶-非核糖体多肽合成酶编码基因成簇存在。这表明该分支的类 SidC 非核糖体多肽合成酶可能是通过基因复制从铁载体合成酶进化而来, 并获得了不同的功能。分支 10~11 的双模块非核糖体多肽合成酶与人类病原菌 *A. fumigatus* 的 SidE 相似<sup>[42]</sup>。SidE 尽管与 NPS2/SidC 铁载体合成酶有着密切的系统发育关系, 但其产物不是铁载体, 而是反丁烯二酰丙氨酸, 其结构与已知具有免疫调节活性的药物十分相似<sup>[42]</sup>。在球孢白僵菌中, 类 SidE 编码基因 BBA\_07589 在感染小菜蛾期间上调<sup>[27]</sup>。因此, 分支 10~11 的非核糖体多肽合成酶可能产生类似反丁烯二酰丙氨酸的产物, 用于抵抗宿主免疫系统的攻击。

分支 16~20 的腺苷酰化结构域来自于单模块非核糖体多肽合成酶。其中, 分支 16 主要存在于白僵菌/冬虫夏草属真菌的基因组中, 它们的编码基因与羧肽酶、双加氧酶、铁 (II) 加氧酶、酰胺水解酶、磷酸酯酶和转运体蛋白等基因共同组

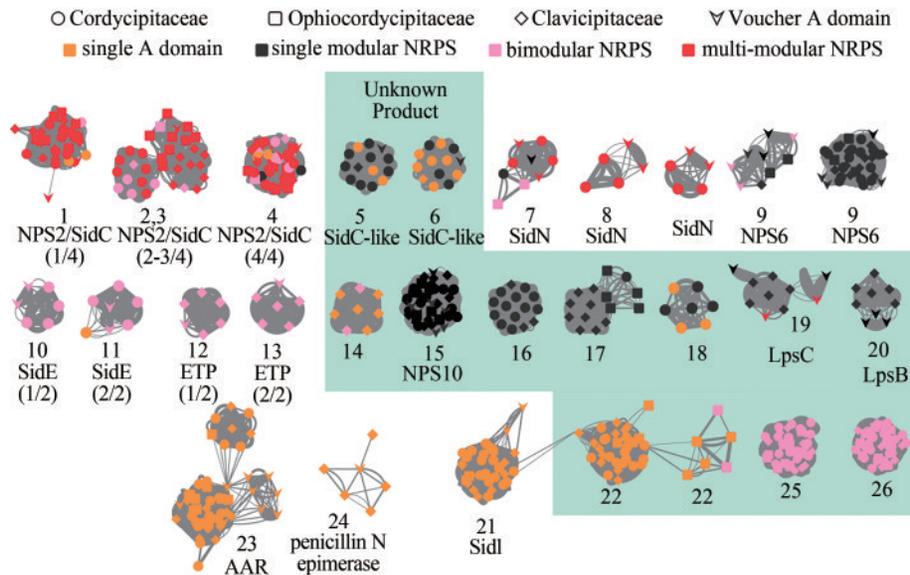


图3 肉座菌目虫生真菌单模块和双模块非核糖体多肽合成酶的相似性网络图

[其中每个节点代表一个腺苷酰化结构域，连接线粗细与相似性成正比。节点颜色显示非核糖体多肽合成酶的类型，形状表示肉座菌目虫生真菌的种类（科），已知功能的结构域用箭头表示。绿色阴影标注产物未知的非核糖体多肽合成酶腺苷酰化结构域。括号中的数字表示腺苷酰化结构域在标签非核糖体多肽合成酶中的位置，例如（1/3）表示该结构域是一个具有3个模块的非核糖体多肽合成酶的第1个腺苷酰化结构域。AAR—L-氨基己二酰半醛缩酶<sup>[24]</sup>；LpsC和LpsB—D-麦角酰胺合成酶亚基1和2，参与麦角生物碱的生物合成<sup>[32]</sup>；ChNPS10—参与真菌形态发育<sup>[33-34]</sup>；NPS2/SidC<sup>[35-36]</sup>和SidI<sup>[37]</sup>—胞内铁载体合成酶；SidN<sup>[38]</sup>和NPS6<sup>[39]</sup>：胞外铁载体合成酶]

Fig. 3 Overview of the A domain distance network for monomodular, bimodular or siderophore-like NRPSs in Hypocrealean entomopathogenic fungal species

(Nodes in the network represent A domains, and the line boldness is proportional to the identity. Nodes are colored according to the domain composition of NRPS. Referenced A domains from voucher NRPSs with known products are indicated by arrowheads. The predicted products/functions of the enzymes are listed below the clades. NRPSs that contain the voucher A domains include: AAR, L-aminoadipate-semialdehyde dehydrogenase<sup>[24]</sup>; LpsC and LpsB: D-lysergyl-peptide synthetase subunit 1 and 2, respectively, involved in ergot alkaloid biosynthesis<sup>[32]</sup>; ChNPS10<sup>[33-34]</sup>: an NRPS-like protein involved in morphological development; NPS2/SidC<sup>[35-36]</sup> and SidI<sup>[37]</sup>: synthetases for intracellular siderophores; SidN<sup>[38]</sup> and NPS6<sup>[39]</sup>: synthetases for extracellular siderophores. Numbers in parentheses indicate the position of the A domain in the voucher NRPS, for example (1/3) indicates the first A domain out of the three housed in the NRPS.)

成基因簇，其中一些基因簇还编码聚酮合酶-非核糖体多肽合成酶杂合酶、酮基还原酶、烯基还原酶、甲基转移酶和细胞色素P450蛋白。分支17中的腺苷酰化结构域主要存在于绿僵菌中，它们的基因簇含有一个寡肽转运蛋白基因，但没有修饰酶基因。分支19~20中的腺苷酰化结构域存在于大多数已测序的绿僵菌属真菌中，与LpsC或LpsB型麦角生物碱非核糖体多肽合成酶十分相似<sup>[32]</sup>，其基因簇中包含萜烯合成酶，可能产生麦角酸 $\alpha$ -羟乙基酰胺型或麦角碱类产物<sup>[43]</sup>。

## 4 结论与讨论

肉座菌目虫生真菌是一个极有特色的真菌类

群，它们在不同的情况下从植物病原菌和腐生真菌进化成为昆虫病原菌，并发展出多种侵染昆虫、抑制其免疫系统、利用昆虫作为自身营养源的机制。同时，许多肉座菌目虫生真菌还保持着与植物共生以及营腐生生活的能力，研究它们的生物合成基因簇不仅有助于理解天然产物在生物和非生物相互作用中的功能，也有助于挖掘临床、兽用或农用药物分子。

近年来基因组测序成本持续降低，基因组分析工具不断完善，公共数据库中高质量真菌基因组序列数量激增，这些极大改变了基因和基因簇功能研究的模式。利用简并引物PCR、黏粒基因文库和类似的传统方法迅速被基于组学方法的基因功能研究所取代，如产生活性代谢产物的真菌菌株的基因组草图测序与分析、基于目标产物结

构的逆生物合成分析、针对产生类似产物菌株的比较基因组分析以及在产生和不产生目标代谢产物的条件下的转录组分析等。类似的基因组学方法已经成为挖掘活性代谢产物及其合成基因簇的主要手段。

越来越多的基因组序列也使得人们发现了大量产物未知的生物合成基因簇。利用异源表达方法,将基因簇转入较为成熟的底盘细胞中激活并表达,逐渐成为天然产物挖掘以及生物合成途径研究的主流方法之一。然而,这些方法仍然面临着重复发现已知产物的问题。因此,尽可能准确地预测这些未知基因簇的功能、筛选最具潜力的研究对象是十分有必要的。对未知基因和基因簇进行聚类分析可以:①发现与合成已知活性产物的基因簇相似的基因簇,从而得到活性产物的类似物,例如分支75~76、分支27~32的合成酶产生具有杀虫、抗菌、细胞毒性等活性的白僵菌素/白僵菌环四肽、破坏素类化合物,分支10~11的合成酶产生可能具有免疫调节活性的类似反丁烯二酰丙氨酸的化合物;②挖掘全新的生物合成基因簇,得到新结构、新功能的化合物,例如分支52~64的合成酶可合成4~8个氨基酸单体组成的环肽,这些合成酶与已知功能的非核糖体多肽合成酶的序列差异较大,且仅发现于肉座菌目虫生真菌基因组中。由于大部分基因在实验室常规培养条件下不表达,因此,研究其产物需要进行异源表达、调控因子激活/抑制或其他转录激活方法。此外,在利用合成生物学方法创制新结构人工产物方面,可根据聚类分析的结果,选取特定功能的基因或基因簇,异源表达部分基因簇、或来源于不同基因簇的基因(即组合生物合成)从而改造已知化合物。

通过对腺苷酰化结构域的聚类来预测真菌非核糖体多肽合成酶的功能也存在一定的局限性。即使在系统发育上接近且整体氨基酸序列一致性很高的腺苷酰化结构域之间,其底物范围也有可能显著不同,如在酶的活性部位的关键氨基酸位点发生改变,就可能改变被激活的氨基酸单体的范围。与原核生物相比,真菌腺苷酰化结构域的底物选择性预测还不够精确,这是因为现有的预测算法往往是根据细菌来源的非核糖体多肽合成

酶构建和训练的。因此,为了更好地建立蛋白序列与产物结构之间的映射关系、从而实现更加准确的功能预测,需要更丰富完善的数据库和更加先进的分析方法。

## 参 考 文 献

- [1] LÜ H N, LIU H W, KELLER N P, et al. Harnessing diverse transcriptional regulators for natural product discovery in fungi[J]. *Natural Product Reports*, 2020, 37(1): 6-16.
- [2] WANG H, FEWER D P, HOLM L, et al. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes[J]. *Proceedings of the National Academy of the Sciences of the United States of America*, 2014, 111(25): 9259-9264.
- [3] MOLNÁR I, GIBSON D M, KRASNOFF S B. Secondary metabolites from entomopathogenic Hypocrealean fungi[J]. *Natural Product Reports*, 2010, 27(9): 1241-1275.
- [4] GIBSON D M, DONZELLI B G G, KRASNOFF S B, et al. Discovering the secondary metabolite potential encoded within entomopathogenic fungi[J]. *Natural Product Reports*, 2014, 31(10): 1287-1305.
- [5] ZHANG X, WEI W, TAN R X. Symbionts, a promising source of bioactive natural products[J]. *Science China Chemistry*, 2015, 58(7): 1097-1109.
- [6] BEEMELMANNS C, GUO H J, RISCHER M, et al. Natural products from microbes associated with insects[J]. *Beilstein Journal of Organic Chemistry*, 2016, 12: 314-327.
- [7] PEDRINI N. Molecular interactions between entomopathogenic fungi (Hypocreales) and their insect host: perspectives from stressful cuticle and hemolymph battlefields and the potential of dual RNA sequencing for future studies[J]. *Fungal Biology*, 2018, 122(6): 538-545.
- [8] OLATUNJI O J, TANG J, TOLA A, et al. The genus *Cordyceps*: an extensive review of its traditional uses, phytochemistry and pharmacology[J]. *Fitoterapia*, 2018, 129: 293-316.
- [9] WANG B, KANG Q J, LU Y Z, et al. Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species[J]. *Proceedings of the National Academy of the Sciences of the United States of America*, 2012, 109(4): 1287-1292.
- [10] DONZELLI B G G, KRASNOFF S B, YONG S M, et al. Genetic basis of destruxin production in the entomopathogen *Metarhizium robertsii*[J]. *Current Genetics*, 2012, 58(2): 105-116.
- [11] XU Y Q, OROZCO R, WIJERATNE E M K, et al. Biosynthesis of the cyclooligomer depsipeptide beauvericin, a virulence factor of the entomopathogenic fungus *Beauveria bassiana*[J]. *Chemistry & Biology*, 2008, 15(9): 898-907.

- [12] XU Y Q, OROZCO R, WIJERATNE E M K, et al. Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of *Beauveria bassiana*[J]. *Fungal Genetics and Biology*, 2009, 46(5): 353-364.
- [13] MOON Y S, DONZELLI B G, KRASNOFF S B, et al. Agrobacterium-mediated disruption of a nonribosomal peptide synthetase gene in the invertebrate pathogen *Metarhizium anisopliae* reveals a peptide spore factor[J]. *Applied and Environmental Microbiology*, 2008, 74(14): 4366-4380.
- [14] SBARAINI N, GUEDES R L M, ANDREIS F C, et al. Secondary metabolite gene clusters in the entomopathogen fungus *Metarhizium anisopliae*: genome identification and patterns of expression in a cuticle infection model[J]. *BMC Genomics*, 2016, 17(8): 399-417.
- [15] DE BEKKER C, OHM R A, LORETO R G, et al. Gene expression during zombie ant biting behavior reflects the complexity underlying fungal parasitic behavioral manipulation[J]. *BMC Genomics*, 2015, 16: 620.
- [16] LIN H Z, LÜ H N, ZHOU S, et al. Deletion of a global regulator LaeB leads to the discovery of novel polyketides in *Aspergillus nidulans*[J]. *Organic & Biomolecular Chemistry*, 2018, 16(27): 4973-4976.
- [17] RUTLEDGE P J, CHALLIS G L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters[J]. *Nature Reviews Microbiology*, 2015, 13(8): 509-523.
- [18] COPP J N, ANDERSON D W, AKIVA E, et al. Exploring the sequence, function, and evolutionary space of protein superfamilies using sequence similarity networks and phylogenetic reconstructions[J]. *Methods in Enzymology*, 2019, 620: 315-347.
- [19] COPP J N, AKIVA E, BABBITT P C, et al. Revealing unexplored sequence-function space using sequence similarity networks[J]. *Biochemistry*, 2018, 57(31): 4651-4662.
- [20] AKIVA E, BROWN S, ALMONACID D E, et al. The structure-function linkage database[J]. *Nucleic Acids Research*, 2014, 42: D521-D530.
- [21] NIELSEN J C, GRIJSEELS S, PRIGENT S, et al. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species[J]. *Nature Microbiology*, 2017, 2: 17044.
- [22] ZIEMERT N, LECHNER A, WIETZ M, et al. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(12): E1130-E1139.
- [23] GALLO A, FERRARA M, PERRONE G. Phylogenetic study of polyketide synthases and nonribosomal peptide synthetases involved in the biosynthesis of mycotoxins[J]. *Toxins*, 2013, 5(4): 717-742.
- [24] BUSHLEY K E, TURGEON B G. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships[J]. *BMC Ecology and Evolution*, 2010, 10: 26.
- [25] KROKEN S, GLASS N L, TAYLOR J W, et al. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(26): 15670-15675.
- [26] RAUSCH C, HOOF I, WEBER T, et al. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution[J]. *BMC Evolutionary Biology*, 2007, 7: 78.
- [27] LIU H, XIE L N, WANG J, et al. The stress-responsive and host-oriented role of nonribosomal peptide synthetases in an entomopathogenic fungus, *Beauveria bassiana*[J]. *Journal of Microbiology and Biotechnology*, 2017, 27(3): 439-449.
- [28] AGRAWAL Y, NARWANI T, SUBRAMANIAN S. Genome sequence and comparative analysis of clavicipitaceous insect-pathogenic fungus *Aschersonia badia* with *Metarhizium* spp[J]. *BMC Genomics*, 2016, 17: 367.
- [29] YANG X Q, FENG P, YIN Y, et al. Cyclosporine biosynthesis in *Tolypocladium inflatum* benefits fungal adaptation to the environment[J]. *MBio*, 2018, 9 (5): e01211-18.
- [30] WIEST A, GRZEGORSKI D, XU B W, et al. Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase[J]. *Journal of Biological Chemistry*, 2002, 277 (23): 20862-20868.
- [31] XU Y Q, ZHAN J X, WIJERATNE E M K, et al. Cytotoxic and antihaptotactic beauvericin analogues from precursor-directed biosynthesis with the insect pathogen *Beauveria bassiana* ATCC 7159[J]. *Journal of Natural Products*, 2007, 70(9): 1467-1471.
- [32] FLEETWOOD D J, SCOTT B, LANE G A, et al. A complex ergovaline gene cluster in epichloe endophytes of grasses[J]. *Applied and Environmental Microbiology*, 2007, 73(8): 2571-2579.
- [33] TURGEON B G, OIDE S, BUSHLEY K. Creating and screening *Cochliobolus heterostrophus* non-ribosomal peptide synthetase mutants[J]. *Mycological Research*, 2008, 112(2): 200-206.
- [34] OHM R A, FEAU N, HENRISSAT B, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen dothideomycetes fungi[J]. *PLoS Pathogens*, 2012, 8(12): e1003037.
- [35] GREENSHIELDS D L, LIU G S, FENG J, et al. The siderophore biosynthetic gene *SIDI*, but not the ferroxidase gene *FET3*, is required for full *Fusarium graminearum* virulence[J]. *Molecular Plant Pathology*, 2007, 8(4): 411-421.
- [36] WALLNER A, BLATZER M, SCHRETTL M, et al. Ferricrocin, a siderophore involved in intra- and transcellular iron dis-

- tribution in *Aspergillus fumigatus*[J]. Applied and Environmental Microbiology, 2009, 75 (12): 4194-4196.
- [37] YASMIN S, ALCAZAR-FUOLI L, GRÜNDLINGER M, et al. Mevalonate governs interdependency of ergosterol and siderophore biosyntheses in the fungal pathogen *Aspergillus fumigatus*[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(8): E497-E504.
- [38] JOHNSON L J, KOULMAN A, CHRISTENSEN M, et al. An extracellular siderophore is required to maintain the mutualistic interaction of *Epichloë festucae* with *Lolium perenne*[J]. PLoS Pathogens, 2013, 9(5): e1003332.
- [39] OIDE S, MOEDER W, KRASNOFF S, et al. *NPS6*, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes[J]. The Plant Cell, 2006, 18(10): 2836-2853.
- [40] BUSHLEY K E, RAJA R, JAISWAL P, et al. The genome of *Tolyocladium inflatum*: Evolution, organization, and expression of the cyclosporin biosynthetic gene cluster[J]. PLoS Genetics, 2013, 9 (6): e1003496.
- [41] DOPSTADT J, NEUBAUER L, TUDZYNSKI P, et al. The epipolythiodiketopiperazine gene cluster in *claviceps purpurea*: Dysfunctional cytochrome P450 enzyme prevents formation of the previously unknown clapurines[J]. PLoS One, 2016, 11(7): e0158945.
- [42] STEINCHEN W, LACKNER G, YASMIN S, et al. Bimodular peptide synthetase SidE produces fumarylalanine in the human pathogen *Aspergillus fumigatus*[J]. Applied and Environmental Microbiology, 2013, 79(21): 6670-6676.
- [43] DONZELLI B G G, KRASNOFF S B, Molecular genetics of

secondary chemistry in *Metarhizium fungi*[M]// LOVETT, LEGER R JST. Genetics and molecular biology of entomopathogenic fungi[M]// Advances in Genetics, 2016, 94: 365-436.



**通讯作者:** 徐玉泉(1971—),男,研究员,博士生导师。研究方向为发现真菌合成的新型生物活性天然产物,阐明其生物合成机制,利用合成生物学的方法对目标产物进行高效生物合成及结构修饰。

E-mail: xuyuquan@caas.cn



**通讯作者:** István Molnár(1963—),男,研究员,博士生导师。研究方向为真菌天然产物生物合成与组合合成生物学。

E-mail: imolnar@arizona.edu



**第一作者:** 张礼文(1985—),女,副研究员。研究方向为从真菌中挖掘具有农用生物活性的次生代谢产物及其合成基因元件,再用于组合生物合成新骨架、改进活性的化合物。

E-mail: zhangliwen@caas.cn