



**HAL**  
open science

## From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites

Thaïs Hautbergue, Emilien L. L Jamin, Laurent Debrauwer, Olivier Puel,  
Isabelle P. Oswald

### ► To cite this version:

Thaïs Hautbergue, Emilien L. L Jamin, Laurent Debrauwer, Olivier Puel, Isabelle P. Oswald. From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites. *Natural Product Reports*, 2018, 35 (2), pp.147-173. 10.1039/c7np00032d . hal-03331311

**HAL Id: hal-03331311**

**<https://ut3-toulouseinp.hal.science/hal-03331311v1>**

Submitted on 1 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# From Genomics to Metabolomics, Moving toward an Integrated Strategy for the Discovery of Fungal Secondary Metabolites

T. Hautbergue<sup>a,b</sup>, E. L. Jamin<sup>a,b</sup>, L. Debrauwer<sup>a,b</sup>, O. Pue<sup>l</sup> and I. P. Oswald<sup>a</sup>

## Abstract

Fungal secondary metabolites are defined by bioactive properties that ensure adaptation of the fungus to its environment. Although some of these natural products are promising sources of new lead compounds especially for the pharmaceutical industry, others pose risks to human and animal health. The identification of secondary metabolites is critical to assessing both the utility and risks of these compounds. Since fungi present biological specificities different from other microorganisms, this review covers the different strategies specifically used in fungal studies to perform this critical identification. Strategies focused on the direct detection of the secondary metabolites are firstly reported. Particularly, advances in high-throughput untargeted metabolomics have led to the generation of large datasets whose exploitation and interpretation generally require bioinformatics tools. Then, the genome-based methods used to study the entire fungal metabolic potential are reported. Transcriptomic and proteomic tools used in the discovery of fungal secondary metabolites are presented as links between genomic methods and metabolomic experiments. Finally, the influence of the culture environment on the synthesis of secondary metabolites by fungi is highlighted as a major factor to consider in research on fungal secondary metabolites. Through this review, we seek to emphasize that the discovery of natural products should integrate all of these valuable tools. Attention is also drawn to emerging technologies that will certainly revolutionize fungal research and to the use of computational tools that are necessary but whose results should be interpreted carefully.

## 1. Introduction

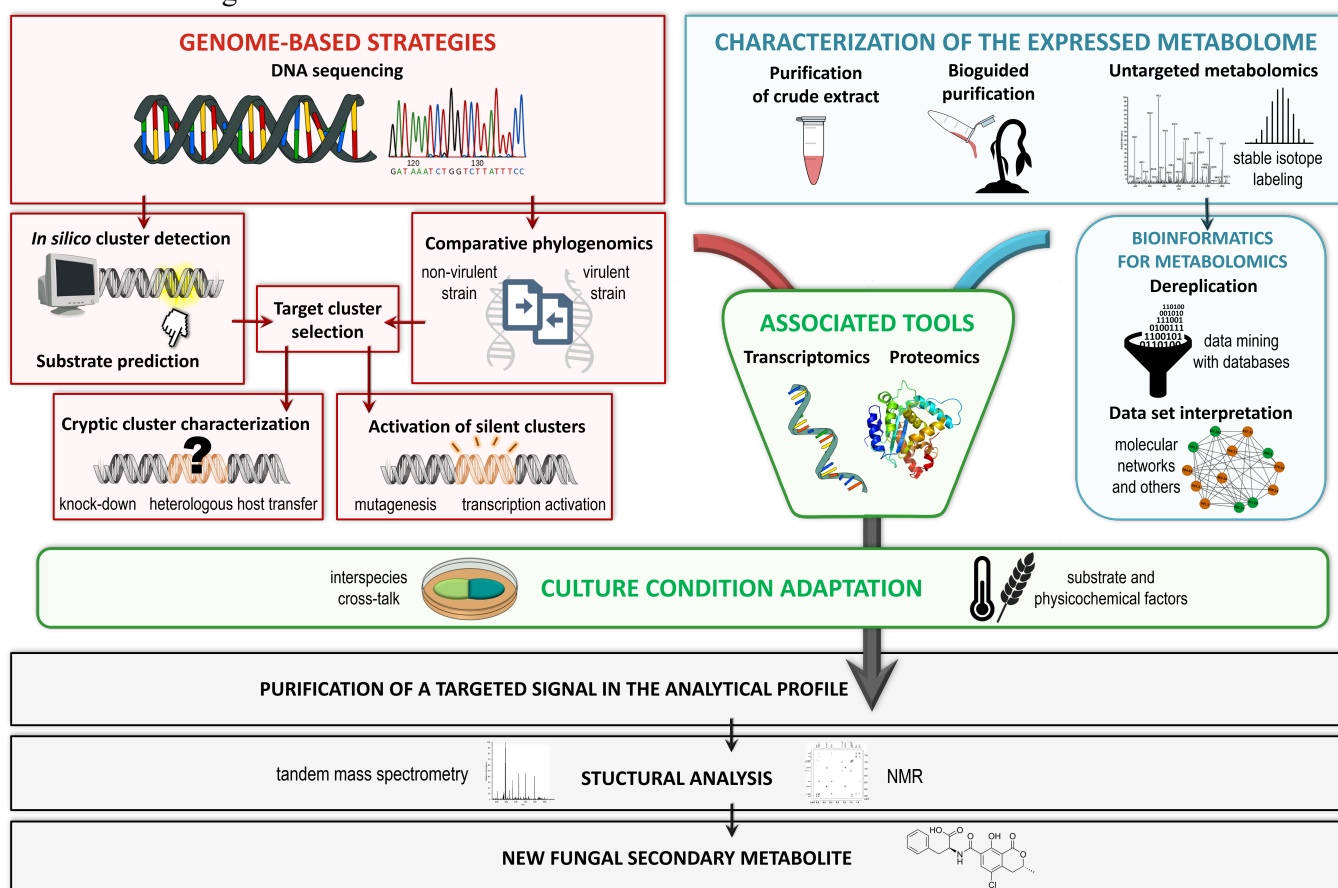
Fungi are well known for their capacity to produce a broad diversity of secondary metabolites that provide them with beneficial properties for adequate growth in a fluctuating environment. On the one hand, these properties are of interest to various industries (particularly pharmaceutical, cosmetic and food industries) that either can commercialize the natural compounds directly or develop derived products from the fungal molecules. Newman and Cragg showed that between 1981 and 2006, 28% of the newly developed industrial chemicals were of natural origin and 24% were inspired by natural products<sup>1</sup>. Over 40% of filamentous fungi are presumed to produce antibiotics under natural growth conditions<sup>2</sup>. Since the discovery of penicillin by Alexander Fleming in 1928<sup>3</sup>, medicine has exploited natural fungal defenses against bacteria to protect humans and animals from pathogenic microorganisms. Likewise, many fungal compounds are used as antimicrobials as well as lipid-lowering medications (lovastatin), immunosuppressants (cyclosporine and mycophenolic acid) and vasoconstrictors (ergometrine).

On the other hand, some fungal secondary metabolites named mycotoxins are the subject of major concern because of their toxicity. Each year, systemic mycoses affecting immunocompromised individuals lead to 1.6-2 million deaths globally<sup>4</sup>. In contrast to mycoses, mycotoxicoses involve intoxication from the exposure to mycotoxins. Ergotism was an epidemic fungal disease recognized in 1676 to result from foods consumption<sup>5</sup>. In the 1960s, the identification of turkey X disease, due to the presence of Aflatoxins represented a turning point in the use of the term mycotoxin<sup>6</sup>, with far more than 400 secondary metabolites being considered as such today<sup>7</sup>. Of these, 30 mycotoxins have been demonstrated to be toxic to humans and/or animals, and only six mycotoxin families are regulated worldwide<sup>8</sup>. Some mycotoxins are also involved in plant diseases as pathogenic or aggressiveness factors, leading to partial or complete destruction of crops and by extension to huge economic losses. The destruction of infected crops was estimated to be equivalent each year to a quantity of food that could feed 600 million individuals, i.e., 8.5% of the world population<sup>9</sup>. Given the environmental concerns seeking to decrease the use of fungicides and the increasing demand for food to feed nine billion people in 2050, the irrepressible development of fungal resistance to pesticides must be faced<sup>4</sup>.

The inherent properties, both beneficial and harmful, of fungal secondary metabolites make the study of these natural products of great importance. Today, 99,000 fungal species are identified and it is estimated that as many as 5 million fungal species exist<sup>10</sup>. Nevertheless, fungal genome investigations of identified species suggest that 80% of their secondary metabolome remains unknown, highlighting the large proportion of compounds waiting to be discovered<sup>11</sup>. Structural

elucidation of unknown fungal secondary metabolites is difficult since many of these natural compounds are synthesized in low amounts in very complex matrices. Each secondary metabolite is secreted under particular environmental conditions to adapt the colonization process, resulting in the production of only a small proportion of the total compounds under standard laboratory growth conditions. Moreover, the toxicity of mycotoxins may be observed after long-term exposure or exposure to mixtures with other fungal compounds. For these reasons, mycotoxin characterization should not be restricted to the predominant metabolites and to the observation of acute biological effects.

This review presents an overview of the large spectrum of methods that have been developed to discover fungal natural products (Figure 1) and presents the advantages of each strategy in addressing the abovementioned difficulties. Strategies focused on the direct detection of the secondary metabolites are firstly reported. Particularly, advances in high-throughput untargeted metabolomics have led to the generation of large datasets whose exploitation and interpretation generally require bioinformatics tools. Some of the promising algorithms for natural product research are presented herein. Then, the genome-based methods used to study the entire fungal metabolic potential are reported. Transcriptomic and proteomic tools used in the discovery of fungal secondary metabolites are presented as links between genomic methods and metabolomic experiments. Finally, the significant influence of the culture environment on the synthesis of secondary metabolites by the fungus is highlighted. Through this review, we seek to emphasize that the discovery of natural products should integrate all of these valuable tools.



**Fig. 1** From genomics to metabolomics, the multiple strategies for the discovery of unknown fungal secondary metabolites.

## 2. Detection of Secondary Metabolites

### 2.1 Targeted Purification

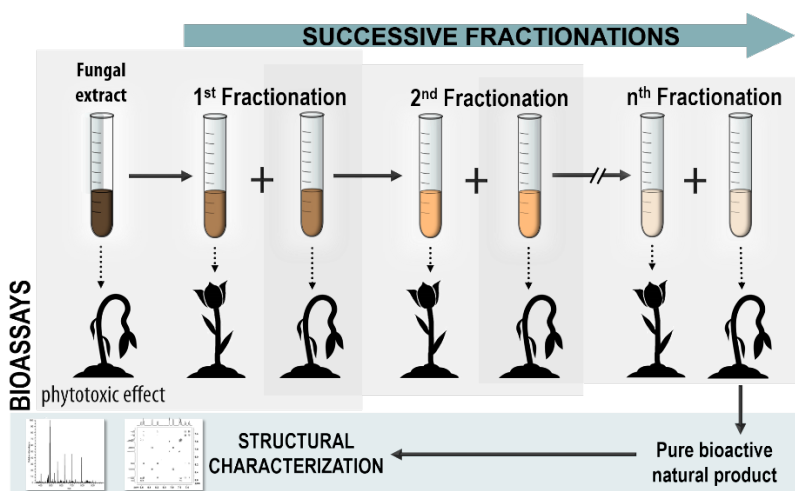
Historically, the detection of fungal secondary metabolites was achieved by two main strategies. One based on the detection of an unidentified signal in a chromatogram, and another one directed by biological effects. They lead to the discovery of thousands of the fungal metabolites and have been widely reported in the literature<sup>12,13,14,15,16</sup>. Here, several examples are presented to illustrate these two strategies discussed hereafter.

**Purification of Crude Extracts.** Many natural products were discovered after observation of an uncharacterized signal in the analytical profile of a fungal extract. Chromatographic purification of the corresponding unknown secondary metabolite and structural analysis by mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy have led to the characterization of many new fungal compounds. Among the numerous examples, two unknown secondary metabolites, arugosins H and G, were isolated from *Emericella nidulans* using normal-phase vacuum liquid chromatography (VLC) and high-performance liquid chromatography (HPLC)<sup>17</sup>. Similarly, two modified peptides were discovered by Garo *et al.*<sup>18</sup>. One of these trichodermamides bears a chlorinated group and displays antimicrobial activities and cytotoxicity against human colon carcinoma. Study of the soil fungus *Trichoderma harzianum* revealed the isoharzianic acid, which is responsible for the inhibition of fungal mycelium radial growth, the stimulation of tomato seed germination and the induction of disease resistance<sup>19</sup>.

Today, the detection of new compounds by simple purification of fungal extracts has been overtaken by numerous advances in analytical technology. Purification of an unidentified signal in the fungal extract can be unsuccessful for several reasons. First, this method may mostly lead to the rediscovery of already known compounds. Second, this strategy is not guided by precise criteria and may lead to identification of compounds without interesting effects. To guide research, a recent study presented the MeHaloCoA algorithm developed to specifically identify halogenated compounds in untargeted MS data from natural extracts<sup>20</sup>. Halogenated natural compounds, such as ochratoxin A, a well-known mycotoxin, and the antifungal griseofulvin, may display biological effects<sup>21</sup>. A successful application of MeHaloCoA led to the discovery of griseophenone I and chlorogriseofulvin, which have antiproliferative effects<sup>20</sup>. This tool may increase the probability of discovering active compounds by simple purification of an unknown signal, although the presence of halogen atoms on the molecule does not guarantee a promising biological effect. Alternatively, extractions guided by the observation of biological effects is another strategy widely referenced.

**Bioguided Purification.** Bioguided purification consists of the consecutive fractionation of a fungal extract that retains the studied effect to ultimately obtain the smallest fraction containing only the active substance (Figure 2). Until today, bioguided research on natural compounds has been performed to trace back to the origin of deleterious effects or of pharmaceutical activities of fungal extracts. The discovery of the penicillin by tracking the antibiotic effect of *Penicillium notatum*<sup>3</sup> is the historical most important example of bioguided discovery of a fungal secondary metabolite. Several toxic fungal secondary metabolites have been identified during sanitary crises, and their characterization was guided by tracking their toxic effects. In 1988, two fumonisins, B1 and B2, which are the cause of leukoencephalomalacia in horses, pulmonary edema in pigs and hepatocarcinogenic in rats, were discovered from *Fusarium verticillioides* cultures using this approach<sup>22</sup>. In an effort to identify anticancer products, Liu *et al.* also used this strategy with *Spicaria elegans* and characterized trichodermamide A, espergillazine A and six cytochalasins<sup>23,24</sup>. Similarly besudon, an antimicrobial, cytotoxic and phytotoxic compound produced by the fruit pathogen fungus *Mollisia benesuada* has been characterized<sup>25</sup>. Recently, this strategy was used to test the antioxidant activity of *Bipolaris sorokiniana*, leading to the discovery of sorokinol<sup>26</sup>.

This approach can be extended to the high-throughput screening of thousands of secondary metabolites to identify compounds with a targeted biological effect. For example, Vansteelandt *et al.* selected a *Penicillium* strain after screening for cytotoxicity against cancer cell lines versus non-tumor cell lines. These activities were found to originate from the synthesis of a chlorinated sesquiterpenoid named ligerin<sup>27</sup>. In searching for molecules capable of inhibiting the SecA protein (involved in bacterial resistance to toxins), Parish *et al.* screened extracts of *Geomyces pannorum* and characterized pannomycin<sup>28</sup>.



**Fig. 2** Principle of the bioguided fractionation strategy to discover secondary metabolites with biological properties: the fungal extract demonstrating bioactivity (here, a phytotoxic effect) is fractionated, and each resulting fraction is tested for the specific activity. The bioactive fraction is then also fractionated, and each resulting subfraction is tested in turn. This protocol is carried forward until the last subfraction contains the pure bioactive compound that can then be characterized (here, by MS and NMR, for example).

Bioguided purification is suitable for secondary metabolites that have strong activities or produced in large quantities but could be limited by the “cocktail effect”. Indeed, this strategy may be unsuccessful in cases where an effect is observed in mixed extract but lost in the purified fraction<sup>11</sup>. Nevertheless, this method is commonly integrated in studies of fungal secondary metabolomes as a complement to other strategies reported in this review<sup>29,30</sup>. For example, bioguided purification could be appropriate to discriminate toxic secondary metabolites among hundreds of compounds highlighted by untargeted metabolomics.

## 2.2 Untargeted Metabolomics

Untargeted metabolomics seeks to obtain a fingerprint of the metabolome of a studied organism using MS or NMR data<sup>31</sup>. Technological advances in MS and NMR methods in recent decades have revolutionized the research on natural compounds by providing high-resolution datasets. At present, this research needs to be directed toward the development of computational algorithms for processing such complex results. Data mining and spectral interpretation remain the critical steps in untargeted research on secondary metabolites<sup>32,33</sup>, particularly when analyzing complex matrices such as fungal extracts<sup>34</sup>. Analytical strategies, dereplication tools and softwares developed to assist untargeted studies of new natural products are described hereafter.

**Stable Isotope Labeling.** One strategy to guide the detection of fungal secondary metabolites in complex metabolomic fingerprints is to specifically label fungal products using stable isotope labeling (SIL). When growing on a labeled substrate, fungi produce labeled secondary metabolites that can be distinguished from contaminants and background noise. Depending on the level of specificity of the labeled substrate, the labeling of fungal secondary metabolites can be restricted to a specific family or spread to the entire secondary metabolome. In their experiment, Klitgaard *et al.* studied the secondary metabolites synthesized from phenylalanine by feeding *A. nidulans* with labeled phenylalanine. The resulting extract was compared with the extract of a non-labeled culture. The compounds of interest were specifically detected using MS by exploiting the given mass difference between the secondary metabolites produced via the incorporation of the <sup>13</sup>C<sub>9</sub><sup>15</sup>N-labeled phenylalanine and the secondary metabolites produced from non-labeled amino acids. Several unknown compounds were identified as several analogues of nidulanin A and fungisporin<sup>35</sup>.

In addition to the precursor-focused SIL strategies, labeling the global source of carbon atoms (<sup>13</sup>C) and/or nitrogen atoms (<sup>15</sup>N) leads to the untargeted labeling of all fungal secondary metabolites<sup>36</sup>. The observation of particular isotopic patterns allows the specific detection of fungal secondary metabolites. Moreover, considering the mass difference between the labeled and unlabeled signals, the number of carbon and/or nitrogen atoms contained in each secondary metabolite can be determined. Bueschl *et al.* cultivated *F. graminearum* on two distinct media, one containing native glucose with almost 100% <sup>12</sup>C and another containing <sup>13</sup>C-glucose as a unique carbon source. Based

on the mass difference between the symmetric isotopic patterns of the labeled and non-labeled co-analyzed metabolites, the carbon atom composition of all the fungal compounds was elucidated and used to determine the chemical formulas<sup>37</sup>. Another approach to the <sup>13</sup>C-labeling of fungal secondary metabolites called isotopic ratio outlier analysis (IROA) is based on the preparation of two labeled metabolomes with specific proportions of labeled nutrients from 5% and 95% <sup>13</sup>C-labeled cultures<sup>38</sup>. Some overviews of IROA methods are available<sup>38,39</sup>. The two strategies based on <sup>13</sup>C labeling of fungal secondary metabolomes are equivalent and the same results are expected. X<sup>13</sup>CMS is a bioinformatics tool that can be used for the processing of raw data from SIL experiments<sup>40</sup>. By allowing the automatic detection of isotopic patterns, this software provides valuable assistance since it decreases the time of data processing compared to manual data mining. However, the false positive rate is likely to be high due to the misinterpretation of background ions. To avoid this, it is highly recommended to check the results obtained by these tools. Consequently, their benefit in terms of saving time is not obvious.

Although calculation of the number of carbon atoms could be sufficient to unambiguously determine the chemical formula of small secondary metabolites (<400 Da), several possible formulas could remain concerning metabolites with high molecular weights, even with measurements of high mass accuracy. To overcome this limitation, Cano *et al.* developed a protocol using double isotope labeling of the fungal substrate with both labeled carbon atoms and labeled nitrogen atoms. This method allows the specific detection of all the fungal compounds as well as the unambiguous determination of their chemical formulas<sup>41</sup>. Three differently labeled wheats were generated after culturing in hermetic chambers alimented with native or <sup>13</sup>C-labeled CO<sub>2</sub> and with native or <sup>15</sup>N-labeled nutrients: (i) a native wheat (containing 99% <sup>12</sup>C), (ii) a <sup>13</sup>C-labeled wheat (containing 97% <sup>13</sup>C) and (iii) a <sup>13</sup>C<sup>15</sup>N-labeled wheat (containing 97% <sup>15</sup>N and 50% <sup>13</sup>C). The harvested wheat grains were then used as only source of carbon and nitrogen atoms for the fungus. After distinct extractions and MS analyses of the three cultures, a list of all secondary metabolites produced by the fungus grown on the wheat grains was generated, and their chemical formulas were unambiguously determined, even for high molecular mass compounds. The second advantage of this method lies in the labeling of the plant with 50% of <sup>13</sup>C and 97% of <sup>15</sup>N, which enables specific and fast detection of each fungal secondary metabolite in the complex mass spectra based on the specific isotopic patterns. This method was successfully validated for the well-documented secondary metabolism of *A. fumigatus*, and a new member of the fumigaclavine family was identified<sup>41</sup>. Similarly, the application of double SIL of a fungal substrate led to the detection of 98 secondary metabolites in *P. verrucosum* and the determination of their chemical formulas. Dereplication of these formulas highlighted that 82 compounds were unknown<sup>42</sup>.

**Dereplication.** As untargeted metabolomics often leads to complex fingerprints of fungal secondary metabolomes, dereplication, i.e., the annotation of known compounds in a mixture, is a key step<sup>43,44,45</sup>. Particularly in the field of secondary metabolite research, Bills *et al.* stated that “dereplication is a vital step in natural product discovery processes”<sup>11</sup>. To avoid wasting time rediscovering known metabolites<sup>11</sup>, this early data-mining strategy seeks to highlight known compounds and guide the analysis toward unknowns by submitting the MS or NMR results to databases. Dereplication must be based on databases adapted to the studied organism and to the type of data (UV, NMR or MS spectra, for example). For fungal studies, databases such as Antibase<sup>46</sup> or the Dictionary of Natural Products<sup>47</sup> list known compounds with the associated metadata such as the chemical formula and exact mass. Since many isomeric products may have the same exact mass, i.e., chemical formula, restricting the databases from global natural products to fungal secondary metabolites could limit the number of matching compounds.

Tandem mass spectrometry (MS/MS or MS<sup>n</sup>) coupled with LC provides both chromatographic retention times and fragmentation pattern specific to each detected secondary metabolite. Hence, some databases reference fungal secondary metabolites with their chromatographic and MS or MS/MS profiles. El-Elimat *et al.* recorded ultra-high performance liquid chromatography (UHPLC) retention times and MS, MS/MS, and UV spectra of 170 known fungal secondary metabolites<sup>48</sup>. Similarly, Nielsen *et al.* listed the LC-UV-MS properties of 474 fungal secondary metabolites<sup>49</sup>, and Kildgaard *et al.* presented a database of 1,300 compounds for the dereplication of marine-derived fungal secondary metabolomes<sup>50</sup>. In addition to these homemade databases, the Global Natural Products Social Molecular Networking (GNPS) system is a general tool for the dereplication of natural products. This open-access library shares approximately 220,000 MS/MS spectra representing more than 18,000 natural products from the MassBank<sup>51</sup>, ReSpect<sup>52</sup> and NIST<sup>53</sup> databases as well as reference compounds from approved worldwide contributors<sup>54,55</sup>. Recently, the dereplication algorithm DEREPLICATOR was associated with the GNPS system<sup>56</sup>. This tool, which

specializes in the dereplication of peptidic natural products from metabolomic experiments, was created to complete the spectral alignment algorithms for the non-ribosomal peptide (NRP) dereplication of cyclic peptides<sup>57</sup> and the platform for the classical dereplication of NRPs, iSNAP<sup>58</sup>. While such databases increase the dereplication efficiency, full identification is only possible from comparison (of the chromatographic pattern and MS/MS, NMR or UV spectrum, for example) with a standard compound.

In addition to the tens of thousands of MS/MS spectra of known compounds, the GNPS system, as the largest library of publicly shared datasets, containing more than 80 million public, uncharacterized, MS/MS spectra, facilitates unknown compound dereplication. Presented as a data-mediated social network, public datasets deposited by laboratories across the globe can be compared with each other. Thus, unknown secondary metabolites identified in different studies can be linked and knowledge can be shared in the interest of natural product characterization<sup>55</sup>.

**Computational Approaches to Assist Data-Mining in Metabolomics.** Advances in metabolomic research on natural products has produced large datasets with hundreds of MS and MS/MS spectra per experiment, leading to the development of computational tools that can help in data mining such as XCMS<sup>59</sup>. Particularly, some algorithms have been developed to help for the interpretation of data generated in a context of natural product research. GNPS enables the organization of hundreds of MS/MS spectra according to their similarity, assuming that their similar MS/MS spectra may originate from similar structures<sup>60</sup>. This organization highlights metabolites with potentially similar structures, including knowns and unknowns, and therefore assists in their structural elucidation.

GNPS workflow has recently been improved by integrating a preprocessing of metabolomic data by the software package for metabolomics MzMine<sup>61</sup>. Mohimani and Pevzner focused on the application of molecular networks for peptidic natural products. Since these secondary metabolites are primarily produced as group of analogs, GNPS simplifies their detection by clustering their similar MS/MS spectra<sup>62</sup>. However, this tool presents some limits. Firstly, by taking into account all neutral losses, specific or not, with the same importance, molecules sharing specific sub-structures may be indistinguishable from molecules sharing non-specific chemical groups. Secondly, the similarity between two MS/MS spectra is calculated via a cosine score giving equal importance to the  $m/z$  and to the intensity of product ions. However, the  $m/z$  should be considered with a higher importance since intensity of product ions is much less representative of the similarity between two molecules. In this context, MS2LDA appears to be a good alternative and/or complementary to GNPS to get further information about similarity between MS/MS spectra<sup>63</sup>. In MS2LDA, an algorithm originally used for text-mining is derived to highlight common groups of neutral losses and product ions, called Mass2Motifs, that may be representative of common substructures.

In addition to these data mining tools, software has been developed to support the interpretation of fragmentation mass spectra and the structural characterization of metabolites. Fragmentation trees (FTs) are data structures generated using particular computational tools to interpret or compare MS/MS spectra<sup>64,65</sup>. Each node in an FT represents an ion (precursor or product ion), and the edges illustrate the fragmentation reactions. Such an algorithm was first designed by Rasche *et al.* based on known fragmentation rules<sup>66</sup> but it was limited by the tremendous number of rules and the fact that totally unknown compounds could fragment according to unknown rules. To address this limitation, Rasche *et al.* created an FT alignment algorithm that compares standard FTs with uncharacterized FTs to highlight the structural similarities between known and unknown natural products<sup>67</sup>. Pairwise alignment similarity matrices of standard FTs showed good correlation with the structural similarities, validating this method. Therefore, this completely automated and “rule-free” analysis is considered a guide to highlight the structural similarities between standards and unknown secondary metabolites. The main limitations of FTs lie in the generation of MS<sup>n</sup> experiments from low abundances of parent ions, particularly in the context of natural product analyses. Caution must also be used because the fragmentation pattern of a molecule depends on the mass spectrometer and fragmentation parameters used. While the GNPS software compares the fragmentation patterns generated from a single analysis or from analyses performed using the same mass spectrometer, FT alignment compares unknown FTs with standard FTs from databases that were generated using different equipment. Users should ensure to not compare the incomparable.

In terms of the structural elucidation of peptidic natural products, classical proteomic tools for peptide sequencing can fail due to the incorporation of non-proteinogenic amino acids and the recurring important structural modifications. To overcome these limitations, NRPquest was created

to compare the MS/MS spectra of multiple peptidic natural products and was successfully used for the *de novo* sequencing of peptides<sup>68</sup>. In addition, the freely available open-source software mMass includes a specific algorithm for the *in silico* fragmentation of peptidic natural products to assist the interpretation of mass spectra of linear and cyclic peptides<sup>69</sup>. Although these tools may provide valuable assistance, predictions from all these computational tools need to be carefully considered. Some experts advice to validate the results by manual interpretation of the MS/MS spectra of interest<sup>70</sup>. Some other bioinformatics tools, such as MassTRIX<sup>71</sup>, CSI:FingerID<sup>72</sup>, CFM-ID<sup>73,74</sup>, MAGMa<sup>75</sup>, MetFrag<sup>76</sup> and MassFrontier, have been developed to support the structural elucidation of unknown compounds, but not specifically natural products.

### 2.3. Critical Insight about the Direct Detection of Secondary Metabolites

Among the different strategies dedicated to the direct detection of fungal secondary metabolites, some historical methods have been outdated and untargeted metabolomics stands as the most promising one. Although mass spectrometry allows the detection of secondary metabolites produced at trace level, the sensibility of mass spectrometers may be not sufficient for their structural elucidation. The development of new technologies such as microflow NMR may be coupled to untargeted mass spectrometry in order to overcome some of the limits related to structural elucidation of new compounds<sup>77,78</sup>. Further improvements of the sensibility of NMR would enable field analyses to characterize secondary metabolites secreted in natural growth conditions impossible to simulate in laboratory.

A critical step of metabolomic research of fungal metabolites is the dereplication. Significant advances are possible with the enrichment of databases. Since NMR data are not useful for rapid annotation of the metabolites in an extract, sharing MS/MS data about new fungal compounds as well as about unknowns should be encouraged. Another important point lies in the standardization of the generation of MS data. Most of the MS/MS spectra in the databases are generated in Collision-Induced Dissociation. However, this mode provides highly variable spectra depending mainly on the type of analyzer<sup>80</sup>. To be useful, the shared data need to be standardized. The development of bioinformatics tools to process mass spectrometry is still recent and promising advances can be reasonably expected in the coming years. In a recent review, Covington *et al.* reported the various MS-based computational methods used to process metabolomic datasets and highlighted their application in the context of natural product research<sup>79</sup>.

Since they are based on different principles, several of these tools should be used to get complementary information. Nevertheless, the bioinformatics results should always be considered carefully, and have to be checked.

The strategies presented above are based on direct identification of secondary metabolites. However, fungus adapt their secondary metabolomes according to their environment. Several gene clusters may therefore be silent in standard laboratory culture conditions. In order to explore the entire metabolic capacity of fungi, the most appropriate strategies are based on the activation of these silent genes by genome mining.

## 3. Genome Mining for the Discovery of Fungal Natural Products

Besides direct detection, genome analysis may allow the identification of fungal secondary metabolites. Thanks to technological advances in molecular biology, the sequencing of microbial genomes has substantially increased in recent decades. Observation of these fungal genomes has led to the establishment of two general principles: (i) genes involved in the synthesis of secondary metabolites are mostly organized in clusters that are principally located in the non-syntenic regions of subtelomeric extremities of the genomes<sup>81</sup> and (ii) these genes code for core enzymes belonging to well-defined families such as non-ribosomal peptide synthetases (NRPSs), polyketide synthases (PKS) and dimethylallyltryptophan synthases (DMATSs)<sup>82-84</sup>. The observation of cryptic clusters, i.e. gene clusters whose corresponding secondary metabolites are still undiscovered, indicates that more than 80% of the fungal compounds remain to be identified. This unexplored portion is partly explained by the fact that many gene clusters, called silent clusters<sup>85</sup>, are not expressed under standard laboratory culture conditions. The misemployment of the term “cryptic clusters” to characterize silent clusters was exposed by Harald Gross who states that “gene clusters are more of cryptic nature due to their undiscovered characters, but not in a sense of silence”<sup>85</sup>. In this review, the term “cryptic” will be used to refer to all the gene clusters whose corresponding secondary metabolites are currently unknown. The term “silent” will be used to refer to particular cryptic



clusters whose products are unknown because they are not expressed under standard laboratory conditions.

Based on genetic observations, a strategy for fungal secondary metabolite discovery consists of the targeted analysis of cryptic clusters and more specifically of the awakening of silent clusters that are not expressed in laboratory growth conditions. Several reviews report the different genome mining tools employed in natural products research<sup>86-90</sup>.

### 3.1 Gene Cluster Analysis

***In silico* Cluster Detection.** The first step of genome mining strategies consists in the detection of gene clusters involved in the synthesis of secondary metabolites. The mostly rule-based bioinformatics algorithms developed for the untargeted detection of these clusters from sequenced fungal genomes are described in some specific reviews<sup>91,92,93</sup>. Owing to the identification of genes coding for well-conserved protein domains, these approaches are very effective at detecting gene clusters that code for known biosynthetic pathways with high precision<sup>91</sup>. SMURF identifies putative clusters by recognizing the conserved sequences of the three major core enzymes, NRPSs, PKSs and DMATs and then evaluates their adjacent genes to test their inclusion in the cluster<sup>94</sup>. Similarly, ClustScan searches all the PKS, NRPS or hybrid PKS-NRPS sequences in a semi-automatic manner<sup>95</sup>. AntiSMASH was developed based on the same principle but has the advantage of analyzing 44 known classes of biosynthesis gene clusters of secondary metabolites<sup>96,97,98</sup>. Recently, PRISM was developed to identify type II PKSs and a wide range of enzymes involved in tailoring reactions<sup>99</sup>. MultiGeneBlast is a program based on a BLAST implementation that performs homology requests on GenBank database or in homemade databases<sup>100</sup>. The search of homology is performed from a sequence of reference and according to multigene modules indicated by the user. One of the advantages of this tool lies in the fact that gene clusters can be searched with different degrees of specificity. On one hand, by indicating sequences common to many core enzymes, it allows the large screening of other core enzymes. On the other hand, by indicating a sequence of reference involved in the synthesis of a particular family of secondary metabolites, it allows the detection of genes involved in the synthesis of putative new compounds of this family. Finally, Wolf *et al.* created a tool called SMIPS-CASSIS to take into account the possible co-regulation of clusters<sup>101</sup>. SMIPS-CASSIS was developed on the hypothesis that some co-regulated clusters include promoters sharing a same transcription factor binding site. Thus, the cluster prediction is guided by the investigation of areas around the core genes, where the promoters share a common binding site.

To identify clusters without core enzymes, some motif-independent protocols have been developed. Takeda *et al.* created an algorithm called MIPS-CG that is based on comparison of the DNA sequences of two genomes<sup>102</sup>. This tool detects the clusters involved in the secondary metabolome by searching for a similar order of genes and considers their high enrichment in non-syntenic blocks. Since this comparative genomics approach is motif-independent, it allows the detection of gene clusters without core enzymes, such as the kojic acid and oxylipin gene clusters. However, it is not applicable for the detection of clusters specific to a particular genome and it might fail for the detection of short gene clusters (less than five genes). Moreover, since order and orientation of genes in a particular cluster are often modified between distant genomes<sup>103</sup>, the results of such algorithms are pertinent when comparing genomes from phylogenetically related species. Using MIPS-CG, cluster boundaries might not be accurate but this limit may be overcome by the simultaneous use of transcriptomic data. Considering this limitation, MIDASS-M integrates genomic and transcriptomic data to highlight the cooperatively regulated genes that are located in the same region of a genome. Using MIDASS-M, it was highlighted that a cluster was three times longer than previously predicted by SMURF<sup>104</sup>. These two methods of motif-independent cluster detection are reviewed in detail by Umemura *et al.*<sup>105</sup>.

It is important to bear in mind that these bioinformatics tools are predictive and that every cluster hypothesis should be confirmed experimentally. These algorithms could miss important information such as genes intertwined in superclusters<sup>106</sup>. The choice of the software used to detect biosynthetic gene clusters depends on the context of the research of new fungal products and on the type of available data. On one hand, motif-dependent protocols, and more particularly AntiSMASH, are widely used and are very useful for quickly acquiring a general view of an uncharacterized fungal genome. These tools should be used when only genomic information about the studied strains is available. Among these tools, SMURF frequently overestimates the size of clusters if two clusters are adjacent and less than 3814 bp apart. This is the case for the patulin cluster which contains 15

genes in all producing species and is identified by SMURF as a cluster containing 23 genes. Equivalent shortcomings had been noted for AntiSMASH but in the version AntiSMASH 4, the determination of the cluster boundaries has been improved<sup>107</sup>. For this, CASSIS and ClusterFinder algorithms have been implemented in AntiSMASH. Although AntiSMASH is continually improved, SMURF is static and does not provide information about similarities between several clusters and others well characterized. On the other hand, since a large part of biosynthetic enzymes diversity is unknown, performing transcriptomic experiments and sequencing several strains of the same species for comparative genomics strongly increase the quality of the information obtained with motif-independent tools. Finally, some limits of motif-dependent algorithms for the detection of biosynthetic gene clusters may be overcome by comparative phylogenomic experiments.

**Comparative Phylogenomics.** Phylogenomics couples genomics and phylogenetics to predict the role of a gene. This method is based on the theory of timeline evolution and focuses on the distribution of genes on a phylogenetic tree<sup>108,92</sup>. Besides providing insights into the distribution and evolution of gene clusters into the fungal kingdom, comparative phylogenomics has been applied to detect gene clusters involved in the synthesis of bioactive secondary metabolites<sup>109,92</sup>. In their review, Van der Lee *et al.* present the study of pan-genomes as an efficient strategy for the detection of unknown gene clusters<sup>92</sup>.

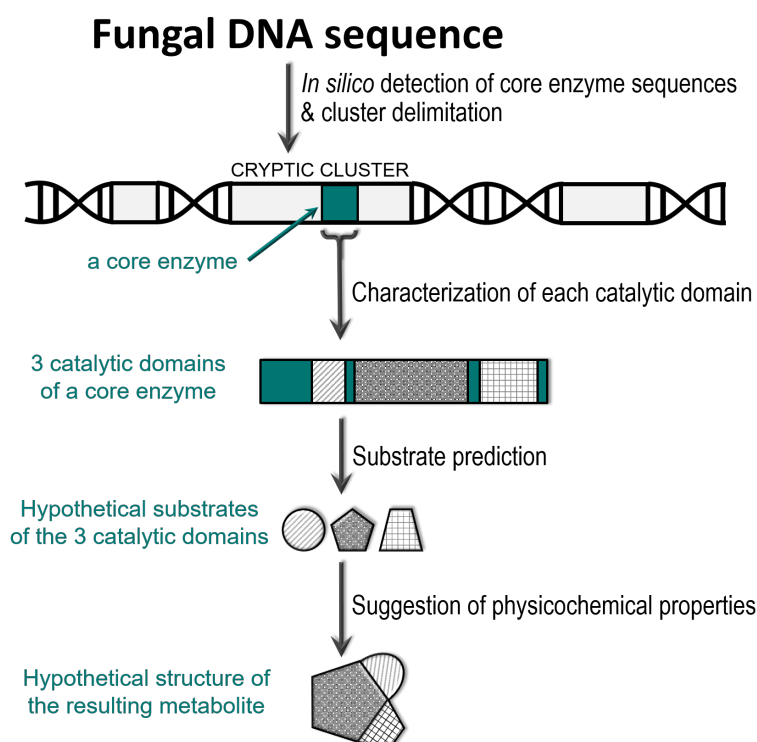
On the one hand, comparative phylogenomics may be applied to link genes and secondary metabolites. In view of this, genome comparison should be restricted to the study of strains within the same species. Indeed, within a given species, the gene clusters involved in the natural products synthesis are diverse but tractable. Thus, the highlighting of common genes as well as genes specific to some strains allows a direct correlation of production of a metabolite with the cluster that encode it. At the scale of a species, comparative phylogenomics may overcome some of the limits of the motif-dependent bioinformatics algorithms for the detection of biosynthetic gene clusters, as well as the frequent wrong genome annotations.

On the other hand, comparative phylogenomics may also be carried out to point out the clusters responsible for a given phenotype. Throckmorton *et al.* focused on the analysis of non-reducing polyketide synthases (NR-PKSs) based on the assumption that the resulting secondary metabolites are virulence factors or chemical defenses for the fungi which synthesize them<sup>110</sup>. By studying the coevolution of the corresponding genes, they identified genes involved in the synthesis of putative toxic metabolites. In an effort to identify genes likely involved in the phytopathogenicity of *Stagonospora nodorum*, Syme *et al.* compared the genomes of three strains: a virulent natural strain, an avirulent strain, and a pathogenic strain that exhibited virulence not observed in the natural strain. This intraspecific genome comparison led to the identification of genes present in the virulent strains but absent in the avirulent strain<sup>111</sup>. Similarly, a phylogenomic comparison of strains of a barley pathogen, *Cochliobolus* spp., revealed clusters specific to particular pathotypes. Deletion of these clusters confirmed their involvement in the virulence<sup>112</sup>.

Studies based on comparative phylogenomics highlighted some gene clusters putatively involved in the synthesis of toxic secondary metabolites. However, this strategy could be inappropriate, particularly in case where there are no adequate comparable genomes. If so, the research of secondary metabolites from a single fungal genome may be guided by the *in silico* prediction of the structure of cryptic cluster's products.

***In silico* Prediction of Cluster Products.** Several algorithms have been developed to detect and identify all the substrate-binding sites in a cluster. Indeed, NRPSs and PKSs include several domains (adenylation domains for NRPSs and acyltransferase domains for PKSs) involved in the specific selection of the building blocks incorporated during the synthesis of the secondary metabolites<sup>113</sup>. These bioinformatics tools may suggest the structural features of the resulting secondary metabolites by predicting the different precursor units involved in their synthesis (Figure 3). Among the substrate predicting tools, the ASMPKS database proposes a polyketide prediction from unknown PKSs<sup>114</sup>. Moreover, the program NP.searcher can consider the tailoring reactions<sup>115</sup>. Other bioinformatics servers reviewed by Anand *et al.*<sup>116</sup> are available, such as SEARCHPKS<sup>117</sup>, NRPS-PKS<sup>118</sup>, SEARCHGTr<sup>119</sup>, PKS/NRPS Analysis<sup>120</sup>, SBSPKS<sup>121</sup>, NRPSpredictor2<sup>122</sup> and NRPSp<sup>123,124</sup>. By increasing the number of reactions and potential substrates considered, the open-source application PRISM improves the accuracy of the structure prediction<sup>99</sup>. A similar approach developed by Minowa *et al.* also includes a program for suggesting the molecular activity<sup>125</sup>. While these tools are able to characterize the different catalytic sites of proteins from their genes, one of their major limitations lies in the prediction of the iteration, that is to say the skipping of catalytic steps during the synthesis of a metabolite by a single enzyme, particularly for PKSs. The above-mentioned tools

predict putative structures from gene clusters in a rule-based way. Actually, the processes through which a secondary metabolite is synthesized, are not totally understood and are likely to follow unknown rules. For these reasons, it is not yet possible to precisely predict the structure of a secondary metabolite from a biosynthetic gene cluster<sup>92</sup>. Therefore, results of these softwares should be considered carefully. One way to consider unknown correlations between enzymes and metabolite structures would be to cross-match thousands of known gene clusters with the structures of the corresponding secondary metabolites. This would be possible for example, by using the MIBiG database developed by Medema *et al.*<sup>126</sup>. Such database, containing gene clusters involved in the synthesis of natural products highlighted by world-wide researchers, could represent a guideline for future efforts to discover new fungal secondary metabolites.

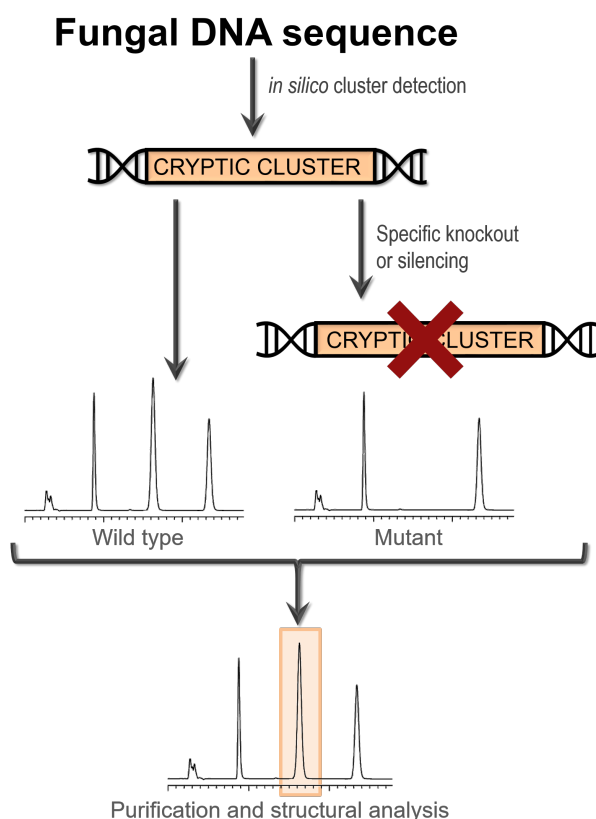


**Fig. 3** Illustration of the bioinformatics strategies to predict NRPS- or PKS substrate specificities and to suggest the structure of the resulting secondary metabolite. After *in silico* detection of the sequence of core enzymes and the delimitation of the cryptic cluster, characterization of the substrate-binding sites of each core enzyme (A-domain for NRPS and AT-domain for PKS) allows the assumption of all building blocks incorporated into the secondary metabolites and, consequently, their physicochemical properties

### 3.2 Knock-out of Target Genes

**Linking Genes and Secondary Metabolites.** Once a cluster of interest is detected, the characterization of its products may be achieved by stopping or reducing its expression. Gene deletions are generally performed to highlight the cluster responsible for the synthesis of a given secondary metabolite. However, this strategy may also be diverted to discover fungal secondary metabolites from cryptic gene clusters. Therefore, a signal that is present in the analytic profile of the wild strain but absent or reduced in the analytic profile of the repressed strain can be attributed to the corresponding secondary metabolite. After locating the target signal, purification and structural characterization of the molecule must be performed from the expressing strain (Figure 4). For example, the four emericellamides C-F were characterized after a double knock-out in *A. nidulans* and a comparison of the resulting LC-MS profiles of the wild strain and of the deleted strain<sup>127</sup>. Knock-out organisms can be created by deletion of the studied sequence via homologous recombination. However, this process is tedious for fungal genomes since fungi are known for a low frequency of well-processed homologous recombination. The incorrect replacement of a target gene by the suitable gene cassette can be partly explained by the non-homologous end-joining pathway that frequently occurs in fungi<sup>128,129</sup>. Inhibition of the Ku70/Ku80 dimeric protein complex involved in repairing DNA double-strand breaks drastically increases the gene targeting efficiency<sup>130,131</sup>. Some recent technologies based on the recruitment of bacterial nucleases were also developed to overcome this limitation. To the best

of our knowledge, no article has reported the discovery of fungal secondary metabolites using these methods, but the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system stands out as a promising tool for knock-out experiments, as discussed in the section 3.6. Besides deletion of cryptic clusters, knock-out experiments may be performed to delete known genes whose products are hiding the signals of uncharacterized secondary metabolites.



**Fig. 4** Specific repression of a cryptic cluster and comparison with the wild-type fungus facilitates locating the corresponding secondary metabolites in the analytical profile. Then, purification and structural analysis steps are performed.

**Deleting Highly Expressed Genes to Decrease Background Noise.** Some clusters may remain uncharacterized for multiple reasons. The difficulty of purifying and structurally analyzing some natural products is a limiting factor, and the production of a major secondary metabolite by a fungus could hide the expression of less abundant metabolites. In some studies, several highly expressed genes were therefore deleted from the fungal genome to decrease the background noise and favor the detection of other compounds. Chiang *et al.* deleted eight of the highly expressed gene clusters in *A. nidulans* and characterized aspercryptin<sup>132</sup>. Similarly, asperfuranone was discovered after deletion of the highly expressed sterigmatocystin gene in *A. nidulans*<sup>133</sup>.

### 3.3 Heterologous Host Transfer

The transfer of cryptic clusters in non-natural hosts represents another strategy for the characterization of cluster products. By cloning an entire gene cluster from a fungus, the host may express the entire secondary metabolite pathway. The difference observed between the wild host and the mutant can therefore be attributed to the expression of the transferred DNA sequence. The ease of establishing a link between a cluster and a natural product once the sequence is expressed by the host is a great advantage<sup>134</sup>. Heterologous host transfer is particularly suitable when the studied fungus is difficult to grow under laboratory conditions or cannot be genetically manipulated<sup>135</sup>. Furthermore, the host genome can be deleted of all of its own clusters before acquiring the metabolite pathway of interest to eliminate interferences and facilitate metabolite identification<sup>83,136</sup>.

*A. nidulans* was presented in several studies as an appropriate host since its metabolome is well characterized<sup>137</sup>. Once its most highly expressed gene clusters had been deleted, this fungus produces only a few secondary metabolites, providing a stable and low background for heterologous expression<sup>136,138</sup>. After *in silico* analyses of dermatophyte fungal genomes, Yin *et al.* observed a cluster of genes similar to that involved in the production of the immunosuppressant neosartoricin.

To investigate the production of metabolites with potential therapeutic interest, the researchers transferred the cluster to *A. nidulans* and discovered neosartoricin B<sup>139</sup>. In another study, a gene similar to a 6-methylsalicylic acid (6-MSA) synthase was observed in the genome of *A. aculeatus*. Heterologous expression in *A. nidulans* combined with substrate labeling showed that <sup>13</sup>C-labeled 6-MSA was incorporated to produce three unknown secondary metabolites named aculins A and B and epi-aculin A<sup>140</sup>. The use of bacterial hosts is another possible strategy for the heterologous expression of fungal genes but presents significant limitations. First, the removal of introns is required but leads to weaker metabolite production<sup>135</sup>. Second, some bacterial hosts such as *Escherichia coli* cannot activate heterologous PKSs and NRPSs using their own intrinsic phosphopantetheinyl (Ppant) transferases<sup>141</sup>. Finally, the broad chemical diversity of fungal PKSs and the consequent requirement of a wide variety of substrates to generate fungal secondary metabolites necessitates the transfer of all the enzymatic machinery involved in the synthesis of such building blocks from the fungus to the bacterial host<sup>141</sup>.

Even if gene transfer to heterologous hosts offers clear benefits and is widely used in bacterial studies, implementing these method could be laborious for fungal genomes<sup>142</sup>. Significant limitations remain, due to both the large size of some clusters (>40 kb), making them impossible to transfer in a single plasmid or BAC vector<sup>143</sup>, and the presence of potential cross-chemistry between clusters on different chromosomes. Moreover, gene transfer often requires exchange between the native promoter and an inducible host promoter<sup>134</sup>. To overcome some of these limitations, Chang *et al.* described a tool that enables the integration of an unlimited number of genes in *A. nidulans*<sup>136</sup>. At each insertion of a gene, a targeted marker is removed and used for the subsequent insertion. Moreover, Bok *et al.* recently developed the first fungal artificial chromosome (FAC) system for the transfer of entire clusters from *A. terreus* to *E. coli* and *A. nidulans*. This process allows the heterologous expression of large DNA fragments (up to 300 kb), large enough to transfer an entire gene cluster with associated regulatory elements. This strategy led to the discovery of terezine D<sup>144</sup>. Recently, the FAC system was integrated in an untargeted metabolism approach allowing a fast and effective screening of thousands of biosynthetic gene clusters<sup>145</sup>. Thus, numerous FAC were generated from different fungal species. Following heterologous host transfer, the resulting extracts were analyzed by mass spectrometry. Then, a pipeline specifically developed to score the data from these analyses allowed highlighting the signals observed only from a particular FAC and therefore resulting from the integration of a gene cluster in the host. By coupling genome mining and metabolomics, Clevenger *et al.* characterized 15 new secondary metabolites and assigned them with their culsters<sup>145</sup>.

### 3.4 Awakening of Silent Gene Clusters by Stimulating Transcription

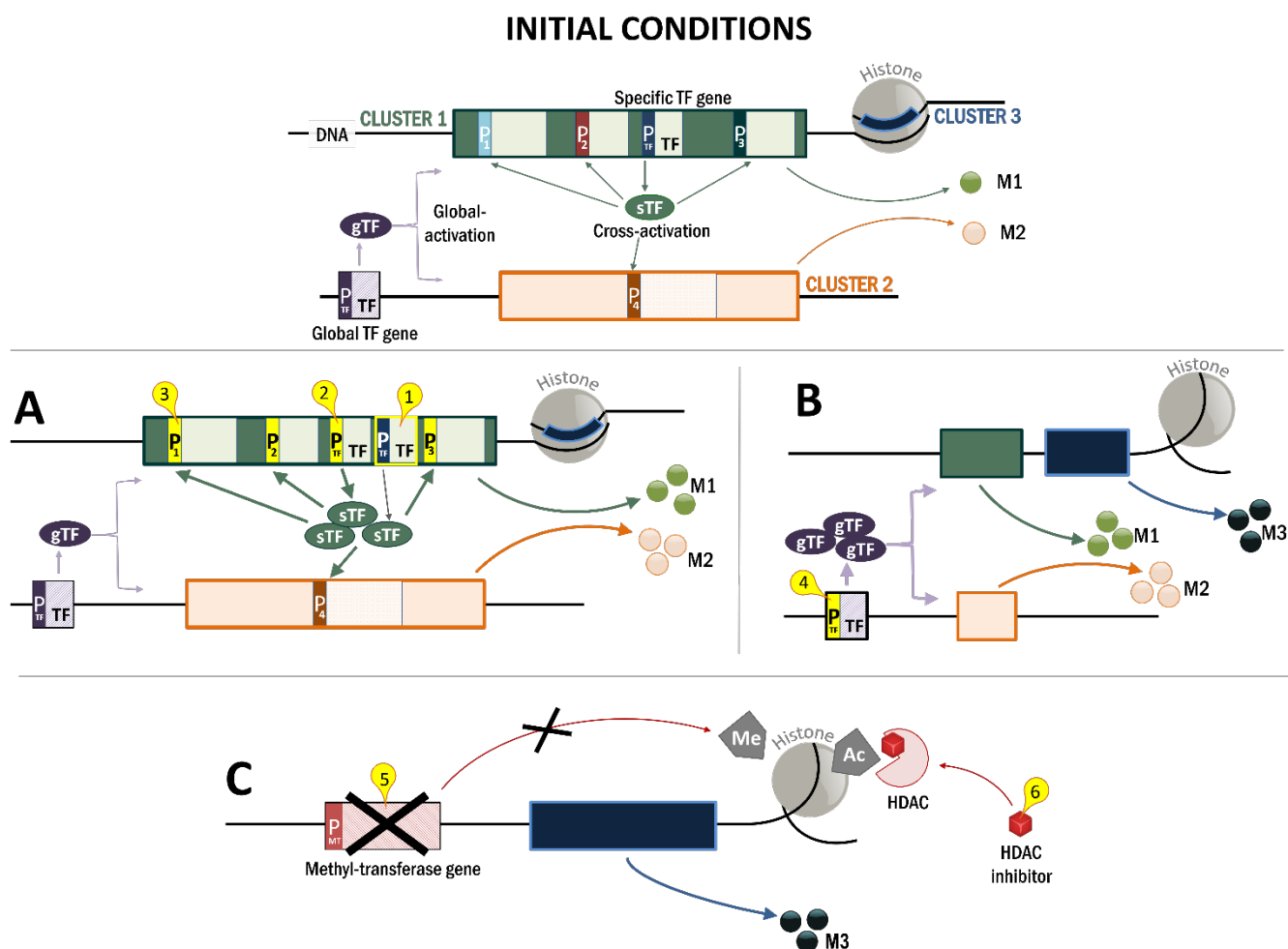
The knock-out and heterologous host transfer strategies were reported as characterization methods for cryptic clusters. In addition, several protocols that focus more specifically on the characterization of clusters not expressed in laboratory growth conditions are intended to awaken silent gene clusters by stimulating their transcription.

**Specific Pathway Overexpression.** Several mechanisms may regulate gene expression in a specific or global pathway. The advantage of the specific modulation of the expression of a gene lies in the fact that it allows the production of the targeted secondary metabolite in significant quantities (for the detection and the structural elucidation) and the establishment of a strong link between the cluster and the natural product<sup>146</sup>. Since a single transcription factor controls the expression of all genes together in a cluster<sup>147</sup>, the overexpression of only one element leads to the transcriptional modulation of the entire target cluster, simplifying cumbersome genetic manipulations. However, this strategy is not always successful as illustrated by articles reporting failures in the transcriptional stimulation of some gene clusters. In searching for all the silent cluster products in *A. nidulans*, Ahuja *et al.* attempted to replace the promoters of all the transcription factors but the experiments failed or the novel compounds were insufficiently produced for characterization<sup>148</sup>. This limitation could be likely explained by posttranscriptional regulation of some specific transcription factors<sup>149</sup>.

- **Stimulation of Specific Transcription Factor Activity.** Many strategies have been developed to stimulate the activity of specific transcription factors. First, some studies multiplied the number of gene copies of the specific transcription factors (Figure 5a1). The efficiency of this method was demonstrated in *A. parasiticus* when Chang *et al.* showed that the inhibited synthesis of aflatoxin can be restored by adding copies of the *afIR* gene, which codes for the specific transcription factor AfIR<sup>150</sup>. Second, specific transcription factors can be overexpressed by coupling their genes with an

inducible promoter (Figure 5a2)<sup>151</sup>. The inducible alcohol dehydrogenase promoter (*alcAp*) is mainly used, but a highly expressed constitutive promoter could also be considered.

Aspyridones A and B were characterized in *A. nidulans* by coupling the *apdR* gene to the inducible promoter *alcAp*<sup>152</sup>. Using this inducible promoter, asperuranone was also characterized in *A. nidulans*<sup>153</sup>. Likewise, three novel fusarielins were characterized in *F. graminearum* by replacing the promoter of the specific transcription factor with the highly expressed constitutive promoter *gpdA*<sup>153</sup>. This latter promoter was also used to activate the expression of a specific regulator in *A. fumigatus* and *Neosartorya fischeri*, leading to the characterization of neosartoricin<sup>154</sup>. Finally, apicidin F, an antimalarial compound, was identified in *F. fujikuroi* after overexpression of the specific regulator ASP2 with the promoter of the  $\beta$ -tubulin gene of *F. graminearum*<sup>155</sup>. Other examples of the discovery of new fungal products by stimulation of specific pathways are presented in specific reviews of genome mining<sup>83</sup>.



**Fig. 5** Strategies for the transcriptional modulation of fungal clusters. The modifications are expressed in comparison with the initial conditions (scheme on top). **A.** Specific transcriptional modulation: Cluster 1 can be specifically overexpressed by (1) multiplication of the number of gene copies of the specific transcription factor (sTF), (2) by the replacement of the promoter ( $P_{TF}$ ) of the specific TF with an inducible promoter or a highly expressed promoter, or (3) by the replacement of the promoters of all the genes involved in the metabolite synthesis ( $P_1$ ,  $P_2$  and  $P_3$ ). Cross-activation may lead to the activation of another cluster (here, cluster 2) by a specific TF. **B.** Global transcriptional regulation: overexpression of a global transcription factor (gTF) by replacement of the promoter ( $P_{TF}$ ) induces a global stimulation of secondary metabolites from distinct clusters (M1, M2 and M3) (4). **C.** Transcriptional modulation by chromatin modification: The expression of clusters that are not expressed in standard culture conditions can be stimulated by (5) the inhibition of histone methylation (Me) via the inhibition of DNA methyl-transferase or (6) by the stimulation of histone acetylation (Ac) via the inhibition of histone deacetylase (HDAC).

• **Promoter Exchange.** Some transcription factors are regulated posttranscriptionally and many clusters involved in the fungal secondary metabolism are not under the control of easily identifiable transcription factors, which is why some studies have focused directly on the activation of the

promoters of genes coding for the different core enzymes (Figure 5a3). In some cases, a single gene determines the expression level of the cluster. However, in fungal genomes, this strategy can be laborious since the replacement of the promoter of all genes is sometimes necessary to stimulate the expression of a cluster<sup>152</sup>. For that reason, this strategy is usually not the first choice for the awakening of silent clusters but is rather used if upregulation of transcription factor failed. Indeed, in the search of the product of the *inp* cryptic cluster, Bergmann *et al.* attempted to overexpress the corresponding specific transcription factor but upregulation of this particular cluster failed<sup>156</sup>. To overcome this limit, Yeh *et al.* replaced the promoter of all genes of the *inp* gene cluster and characterized the new proteasome inhibitor fellutamide B<sup>157</sup>. Similarly, the production of microperforanone by *A. nidulans* has been highlighted by exchanging the native promoter of all 13 silent NRPS-like genes<sup>158</sup>.

**Global Transcriptional Modulation.** In the absence of a specific regulation, silent clusters are awakened by global stimulation of the transcription. Global transcription factors are coded by genes in non-cluster areas and also regulate genes not involved in secondary metabolism (Figure 5b)<sup>147</sup>. The most referenced global regulator in terms of fungal secondary metabolism is LaeA<sup>159</sup>. *A. fumigatus* mutant experiments showed that this factor is involved in the transcriptional stimulation of 40% of the biosynthesis genes of fungal secondary metabolites<sup>160</sup>. Wiemann *et al.* even argued that there is no other known global regulation factor that awakens as many silent clusters as LaeA and its homologous do<sup>83</sup>. However, posttranscriptional modulation of global transcription factor had been reported, as for specific transcription factor, and may lead to the fail of this strategy<sup>159</sup>. By modulating of LaeA in *A. nidulans*, Bok *et al.* characterized terrequinone A, an antitumor compound never identified previously in this fungus<sup>161</sup>. VeA is another global regulator specific to fungi that is involved in secondary metabolite regulation and could also be manipulated to explore natural products<sup>162,163</sup>. While the commonly used regulators such as LaeA are responsible of the upregulation of secondary metabolomes, the gene *mcrA* has just been discovered as a gene likely involved in the global downregulation of fungal metabolism<sup>164</sup>. Indeed, deletion of this gene led to the upregulation of hundreds of genes in *A. nidulans*, even in strains deleted for *LaeA*. Homologous genes have been found at least on *Penicillium* and *Aspergillus*, which offers promising prospects for genome mining and fungal product discovery.

Despite the significant use of global transcriptional modulation for secondary metabolite discovery, the mode of action by which these global transcription factors positively or negatively control metabolite synthesis remains uncertain. Since LaeA was suspected to possess methionine methylation<sup>165</sup> and histone acetylation functions<sup>166</sup>, it is suggested that the transcriptional regulation of cluster expression is regulated by LaeA through chromatin modification.

**Mutagenesis Strategies.** Transcriptional activation of gene clusters that remain silent in standard culture conditions may be achieved, as previously described, by stimulation or restriction of the expression of several actors involved in the transcription of genes. In view of this, mutation of fungi may be performed, for example by deletion of *mcrA*. However, such targeted mutations only concern known transcriptional regulation processes. Since it is likely that little is known about the global system of fungal transcription regulation, other strategies, called mutagenesis strategies, consist on the induction of random mutations in a fungal genome and on the comparison of the metabolic profiles of the wild and the mutant strains. New signals may therefore appear in the signal of the mutant strain and new metabolites may be discovered.

The most widely used mutagenesis strategy for gene cluster awakening consists of inducing random mutations by UV irradiation. One of the first fungal experiments was conducted with *Sphaeropsidales* sp. (strain F-24'707), and UV irradiation led to 25 different mutants. Comparison of the metabolic profiles showed that one of the mutants produced a novel metabolite, mutolide<sup>167</sup>. In addition to UV irradiation, chemical mutagenesis can be performed using several carcinogenic and mutagenic chemicals such as 4-nitroquinoline 1-oxide, known to mimic the biological effect of UV irradiation<sup>168</sup>. Using this 4-nitroquinoline 1-oxide, Ramamoorthy *et al.* studied the regulation of the expression of gene clusters in *A. nidulans* and highlighted the regulation factor *rtfA*<sup>169</sup>.

Ribosome engineering is another mutagenesis strategy. Before being applied to fungal studies, this principle for the activation of silent genes was described in bacteria<sup>90</sup>. The induction of antibiotic resistance, acquired by mutations of ribosomal protein S12 ~~ribosomes~~ (gentamycin resistance) or mutation of RNA polymerase (rifampicin resistance), was correlated to an awakening of silent gene clusters involved in the secondary metabolome<sup>170</sup>. The mutation of RNA polymerase increases its

affinity for promoters and enhance the transcription of genes involved in some silent gene clusters while ribosomal protein mutation enhances protein synthesis<sup>171</sup>. As fungi are not naturally sensitive to antibiotics, antibiotic resistance has been addressed in several different ways. Antibiotics were combined with dimethyl sulfoxide (DMSO) to increase the fungal membrane permeability to these molecules and allow the development of drug resistance in *P. purpurogenum*. The incubation of fungal spores with DMSO and gentamicin or neomycin created mutants with anti-tumor activity and led to the characterization of five secondary metabolites, including the recently identified penicimutide<sup>172,173,174</sup>. In another approach, drug resistance linked to the awakening of silent clusters was induced by incubation of the fungus with diethyl sulfate (DES) to characterize the secondary metabolome of *P. purpurogenum*. Nine antitumor compounds and three rare chromones were discovered<sup>30,175,176,177</sup>. Finally, membrane permeability and neomycin-resistance were induced in *A. versicolor* using an ultrasound-mediated approach, and the induced aminoglycosides resistance was linked to the synthesis of unknown antitumor compounds<sup>178</sup>.

Mutagenesis strategies are based on the hypothesis that random mutations may awaken silent clusters, but the locations of the mutations in the DNA sequences have not been described. It has not been demonstrated that the compounds result only from the transcriptional activation of a cluster. Random mutation may interrupt a biosynthetic pathway, highlighting some of the precursors<sup>179</sup>. Moreover, some mutations may lead to the production of modified enzymes, explaining the degradation of antibiotics on the one hand, and the generation of new compounds by modification of known secondary metabolites on the other hand<sup>180</sup>.

### 3.5 Awakening of Silent Gene Clusters by Chromatin Modification

The chromatin modification of fungal genomes has been considered an efficient way to control gene expression since it was observed in *A. nidulans* that the deletion of *hdaA*, which codes for an HDAC, generates strong transcriptional activation of sterigmatocystin and penicillin genes<sup>181,182</sup>. Two types of strategies can be used to control the chromatin modification factors. First, the use of chemical inhibitors has the advantage of not requiring genetic manipulation (Figure 5c6)<sup>183</sup>. This strategy consists of using different types of chemical chromatin modifiers to modulate histone acetylation, methylation, phosphorylation, ubiquitination, sumoylation, glycosylation, or neddylation<sup>84,134,151</sup>. Second, several genes involved in histone modulation can be overexpressed or repressed (Figure 5c5). Among the different chromatin modification processes, the two that are primarily modulated are histone methylation (i.e., repression of transcription) and histone acetylation (i.e., stimulation of transcription).

**Modification of DNA Methylation.** 5-Azacytidin is a DNA methyltransferase (DNMT) inhibitor commonly used in research on secondary metabolites<sup>134</sup>. The first antidiabetic and anti-inflammatory compound produced by a microorganism, (S)-(+)-sydonol, as well as three analogous compounds were discovered in *A. sydowii* after the addition of this inhibitor<sup>184</sup>. Culturing fungi with this DNMT inhibitor enabled the characterization of oxylipins produced *de novo* in *Cladosporium cladosporioides*; the identification of lunalides A and B in *Diatrype* sp.<sup>185</sup>; and the discovery of two secondary metabolites in *P. citreonigrum*, atlantinones A and B<sup>186</sup>.

In addition to these epigenetic modifier additions in the culture medium, mutants have been created by deletion of natural chromatin modifier genes<sup>187</sup>. Recently, knocking out KMT6, a histone methyltransferase in *F. fujikuroi*, activated the expression of four silent genes coding for core enzymes. After heterologous recombination of the uncharacterized sequence in *E. coli*, a novel sesquiterpene was identified<sup>188</sup>.

**Modification of Histone Acetylation.** Several studies of silent gene clusters focused on the stimulation of their transcription by activation or inhibition of histone acetylation (Figure 5c6). Indeed, it has been demonstrated that some gene clusters were expressed after inhibition of histone desacetylases or activation of histone acetyltransferases such as EsaA<sup>189</sup> and Gcn5<sup>190</sup>. Use of the HDAC inhibitor suberoylanilide hydroxamic acid (SAHA) led to the discovery of cladochromes F and G from *C. cladosporioides*<sup>185</sup> and nygeron A from *A. niger*<sup>191</sup>. Testing different ways to maintain histone acetylation, Albright *et al.* cultivated *A. nidulans* in the presence of diverse HDAC inhibitors (SAHA, trichostatin A or sodium butyrate) and identified a fellutamide and two antibiotics never characterized previously in this fungus<sup>192</sup>.

Another approach to controlling protein acetylation under laboratory conditions involves genetic engineering. In the above study, Albright *et al.* transferred to the fungus an inducible acetylase that remains silent in the absence of stimulation. The secondary metabolomes of the induced and



repressed genomes were compared, and the results confirmed the results from the experiments with the HDAC inhibitors.

### 3.6. Critical Insight about Genome Mining for New Fungal Compounds Characterization

The discovery of secondary metabolites produced by cryptic gene clusters can be performed by genetic engineering, but several limitations remain, including the masking of unknown metabolites by the highly expressed metabolites and the silencing of the studied genes. These problems have been addressed through the development of techniques such as knock-down experiments, gene transfer to heterologous hosts and transcriptional activation of silent gene clusters. Although metabolomic experiments are increasingly used, the genome-mining strategies are currently the most commonly employed strategies for the research of new fungal secondary metabolites. Stimulation of the transcription of gene clusters stands as one of the most promising strategies for the characterization of new fungal products. However, methods that generally require deletion, insertion or mutation of DNA sequences may be laborious in fungi because of their high frequency of non-homologous end joining which induces incorrect insertions of DNA sequences. Recent advances, and particularly the CRISPR-Cas9 system, will certainly increase the use of genome mining for the awakening of gene clusters<sup>193</sup>. Discovered as an adaptive immunity of bacteria against viruses, the CRISPR-Cas9 system involves two components that induce double-strand breaks at specific loci of the DNA strand: a bacterial endonuclease Cas9 and a single-guide RNA (sgRNA) sequence that targets the DNA sequence. Firstly, by simply designing specific sgRNA sequences, gene deletions, insertions or mutations can be precisely and efficiently induced in every organism<sup>194</sup>. In 2016, the first gene mutations, deletions and replacements were performed in fungi using CRISPR-Cas9<sup>195,196,197,198</sup>. This technology has been reported in recent papers about fungi as a very promising tool for specific or global transcriptional modulation<sup>199,200</sup> since it offers unparalleled potential for accurate and high throughput manipulations of fungal genomes. Secondly, one of the major limits of chromatin modification for transcriptional stimulation of fungal metabolomes lay in the fact that these modifications are untargeted. Comparing to targeted mutation, chromatin modification did not allow correlation between a new secondary metabolite and a particular gene. Today, it is however possible to guide chromatin modification by coupling a Cas9 without catalytic activity with a protein involved in the desired modification<sup>201</sup>.

The way by which fungal secondary metabolome is regulated is poorly understood, even in well-known fungi such as *A. nidulans*. Increasing knowledge about the transcriptional regulation mechanisms would allow the development of new strategies for the awakening of silent gene clusters and would improve the research of new fungal products. The recent discovery of the global downregulator *mcrA* illustrates the need for further investigations of the mechanisms at the origin of the regulation of fungal secondary metabolomes.

The great part of fungal secondary metabolites waiting to be discovered has been highlighted thanks to *in silico* recognizing of biosynthetic gene clusters. These bioinformatics tools are crucial to investigate the genome of a studied fungus and are almost systematically used to evaluate its metabolic potential. However, these algorithms present some limits and further improvements are necessary to make them more reliable. *In silico* cluster detection should be performed by taking into account that gene clusters may consist of non-core enzymes on the one hand, and by considering that little is known about the correlation between genetic sequences of enzymes and chemical structures of the corresponding secondary metabolite on the other hand. To consider these limitations, some softwares such as MIDASS-M consider both genomic and transcriptomic data. Indeed, transcriptomic as well as proteomic experiments might provide additional information and help in the characterization of fungal secondary metabolites.

## 4. Transcriptomic and Proteomic Approaches

### 4.1. Transcriptomic approaches

Discovery of natural products can be based on genomic engineering of cryptic clusters or on direct observation of the fungal secondary metabolome, but several studies also used transcriptomics as an intermediate highlighting the links between genome and metabolome. Thus, the different transcriptomic techniques are a way to support one or several of the abovementioned methods rather than fully fledged strategies for the discovery of fungal secondary metabolites. In this context, microarray and RNA-seq (or whole transcriptome shotgun sequencing) methods are applied with two

objectives. First, these methods are useful for highlighting co-regulated gene clusters, i.e., clusters involved separately in the synthesis of different secondary metabolites but whose expressions are linked. For this purpose, the abovementioned bioinformatics tool for cluster detection, MIDASS, integrates an algorithm that highlights co-regulated clusters by comparing the transcriptomic data obtained under metabolite-producing and non-producing conditions<sup>104</sup>. More specifically, transcriptomics may be adapted to detect clusters containing genes located on different chromosomes. Since studies had demonstrated examples of this cross-chemistry between gene clusters<sup>156</sup>, Andersen *et al.* created a microarray compendium specific to *A. nidulans* from 44 culture experiments. Each clustering score was calculated to yield a numerical value with which to correlate the expression profiles of two genes. Using this method, the researchers highlighted a secondary metabolite, nidulanin A, which involves genes from at least two chromosomes<sup>202</sup>. Nevertheless, RNA studies should include the detection of the corresponding secondary metabolites since many posttranscriptional regulations may disrupt the expression of a cluster, such as the posttranscriptional regulation of the synthesis of sterigmatocystin by PkaA in *A. nidulans*<sup>203</sup>.

Second, transcriptomic techniques could be applied to understand the links between genes and secondary metabolites or between genes and active properties. Thus, the clusters suggested by comparative phylogenomics to be responsible for the pathogenicity of *S. nodorum* were found to be specifically expressed during wheat infection based on comparison of *in planta* and *in vitro* microarrays<sup>204,205</sup>. Moreover, RNA-seq was implemented in studies of transcriptomic variations during the different stages of fungal infection<sup>205,206</sup>. An experiment involving a microarray comparison of *in planta* cultures led to the observation of a particular unknown cluster involved in wheat infection by *F. graminearum*<sup>207</sup>.

#### 4.2. Proteomic Approaches

Research of fungal secondary metabolites can be performed by studying and modulating every process involved in the synthesis of these molecules. Although genome-based strategies and metabolomics are the most employed strategies, sometimes coupled with transcriptomic experiments, fungal secondary metabolomes can also be explored from a proteomic perspective. Hereafter, strategies based on the analysis of the enzymes involved in the synthesis of secondary metabolites are presented as complementary approaches. Proteomic approaches are useful to take into account the post-transcriptional modifications that may occur during the synthesis of fungal secondary metabolites. Thus, the studies presented hereafter are based on the assumption that changes in the proteome (in terms of types and quantity of enzymes, such as NRPS, PKS or DMAT for example) reflect changes in the secondary metabolome.

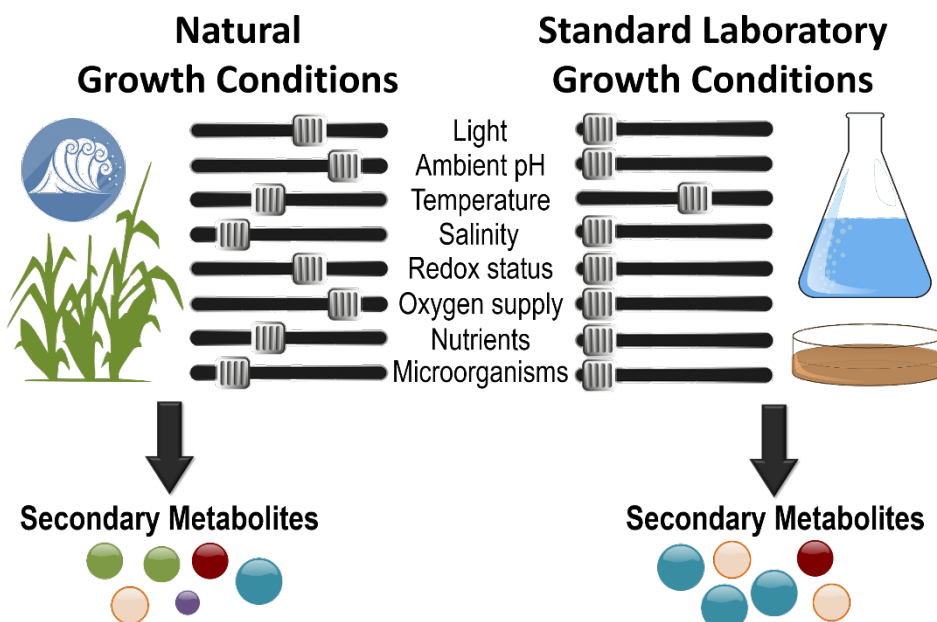
Detection of proteins could be performed using 2D polyacrylamide gel electrophoresis (PAGE)-LC-MS/MS analyses. Thus, Owens *et al.* studied the proteome variations of *A. fumigatus* under different conditions of oxidative stress. The proteins in the crude extract were separated by 2D sodium dodecyl sulfate (SDS)-PAGE. Proteins with quantitative differences between the experiments were excised, digested by trypsin and analyzed by LC-MS/MS. The results demonstrated the presence of approximately 20 enzymes that were involved in the secondary metabolome and specifically increased during oxidative stress<sup>208</sup>. A similar approach, called the Proteomic Investigation of Secondary Metabolism (PrISM), targets the core enzymes of the secondary metabolome by selecting the high-molecular weight proteins after 2D SDS-PAGE separation. After a tryptic digest, NRPSs and PKSs are specifically detected in MS/MS data by tracking the neutral losses corresponding to the labile posttranslational modifications of the specific domains of these enzymes. After characterization of the enzymes, their PCR (polymerase chain reaction) primers are designed, the corresponding gene clusters are identified, and the corresponding secondary metabolites are characterized<sup>209,210</sup>. Of note, gel-based proteomics methods are often restricted to the analysis of abundant proteins<sup>211</sup>.

To study proteins of low abundance, a strategy involving the use of isobaric tags for relative and absolute quantification (iTRAQ) was developed<sup>212</sup>. Concerning fungi, this technique has the advantage of enabling the analysis of proteins secreted in low amounts (for example in the early stage of plant infection) and the quantitative comparison of 4 samples in the same analysis<sup>213</sup> (to compare different culture conditions or different growth stages, for example). This strategy was used by Taylor *et al.* to detect the biosynthetic pathways specifically upregulated by *F. graminearum* when growing in conditions favorable to mycotoxin production<sup>213</sup>. Methods linking stimulation of the transcription of fungal genes clusters by both specific and global regulators and proteomic approaches also allow the identification of NRPS enzymes secreted in low amount<sup>214</sup>. Other methods have been developed to improve the interpretation of complex mass spectra generated from protein extracts but have not been applied in fungal studies<sup>215,216</sup>.

Since they allow considering post-transcriptional modifications, proteomic experiments are sometimes integrated in the research for fungal secondary metabolites. However, recent advances in metabolomics provide an easy way to detect directly the secondary metabolites, also by taking into account post-transcriptional modifications. Moreover, identification of fungal enzymes might be easier by sequencing the genome and using software for automatic detection of the corresponding genes. The alteration of gene expression may also be highlighted by RNA sequencing or gene expression microarrays rather than by detection of enzymes by LC-MS/MS. Although such experiments may couple genomics and transcriptomics or proteomics to carefully consider the alteration of gene expression, some other strategies precisely aim to modify the expression of genes by modifying the fungal growth environment. Here, the purpose is the same as genome-based strategies stimulating silent genes transcription (see section 3.4), but without complex genetic manipulations.

## 5. Modification of the Growth Environment: Environment-Dependent Approaches

A major factor to consider in research on fungal secondary metabolites is the influence of the culture environment. To explain why 80% of the potential fungal secondary metabolites remain unknown, it was suggested that a majority of the gene clusters are silent under standard laboratory culture conditions<sup>83</sup>. Fungi modulate the synthesis of secondary metabolites depending on their needs for defense or colonization of the environment. This principle, commonly called OSMAC (one strain many compounds), indicates that changes in the growth conditions can completely modify fungal secondary metabolism. In a review dedicated to bioactive fungal secondary metabolites, Bills *et al.* even argue that the lack of this knowledge was historically a major limiting factor in the discovery of fungal compounds of industrial interest<sup>11</sup>. Standard laboratory culture conditions are frequently arbitrarily monitored, and only several environmental factors, such as the temperature, nutrient availability or light exposure are controlled (Figure 6, right side). In contrast, natural growth conditions involve a complex balance between numerous variables that may fluctuate independently as well as the co-exposure to other fungi/bacteria (Figure 6, left side). Two different approaches to modify fungal growth conditions are considered: interspecies cross-talk and modification of the physicochemical culture conditions. Hereafter, we describe some studies in which modification of the growth environment was used as the strategy to discover fungal secondary metabolites. However, this method is commonly integrated as a complement to other strategies reported in this review.

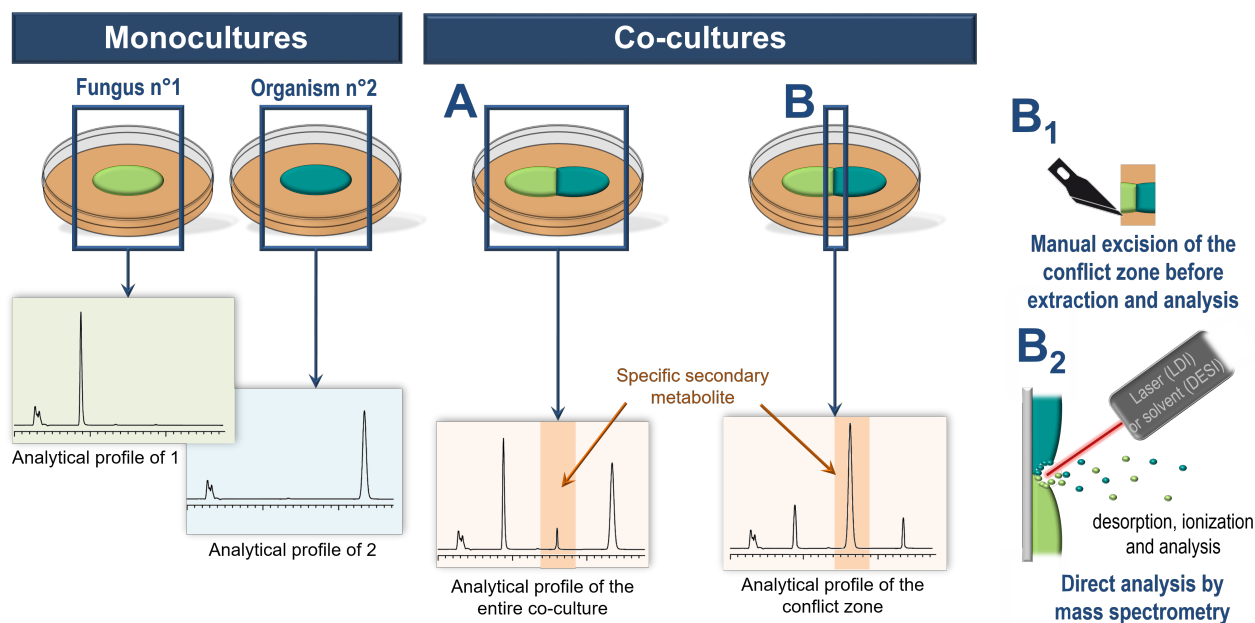


**Fig. 6** Illustration of the differences between natural growth conditions (sea or cereal crops for example) and standard laboratory growth conditions (liquid or solid media) for a fungus in terms of physical, chemical and interspecies cross-talk parameters. The slider positions reflect the complexity of the natural parameters involved in the production of the fungal secondary metabolome on the left side and their basic modulation under laboratory conditions on the right side. These changes in terms of quantity and type of the secondary metabolite depending on the variation in each parameter are expressed under the general term OSMAC (One Strain-MANy Compounds).

## 5.1 Interspecies Cross-Talk

The first approach for changing culture conditions involves the co-culturing of a fungus with other fungi or bacteria<sup>217</sup>. Interspecies cross-talk is used in research on antimicrobial compounds because co-culture seeks to stimulate the production of secondary metabolites involved in fungal defense against other microorganisms<sup>142,147,218</sup>. It was also reported that fungi might communicate with other microorganisms by integrating stimuli, for example, via chromatin modifications<sup>219</sup>. As reported by Marmann *et al.*<sup>220</sup>, these strategies for natural product characterization can be based either on the analysis of the entire mixed fermentation or on the specific observation of the conflict zone.

In some studies, compounds were identified by comparing the analytical profiles of monoculture extracts with that of the entire co-culture extract<sup>221,222</sup>. Thus, a signal present in the co-culture but absent in the monocultures originates from secondary metabolites specifically synthesized in response to the conflict between the two organisms (Figure 7). Zuck *et al.* characterized two xanthocillin analogues in a co-culture of *A. fumigatus* with a bacterial strain, one of which exhibited submicromolar activity against tumor cell lines<sup>223</sup>, and three lipoaminopeptides were characterized<sup>224</sup>. Similarly, a co-culture of *F. tricinctum* and *Bacillus subtilis* led to the identification of two secondary metabolites: macrocarpon C and a derivative of anthranilic acid<sup>225</sup>. König *et al.* studied the influence of *Streptomyces rapamycinus* on the secondary metabolome of *A. fumigatus*. A group of unknown metabolites was highlighted, and fumicyclines A and B were characterized<sup>226</sup>. Through additional knock-out and chromatin modification experiments, the authors have shown that the secretion of these molecules was stimulated by the bacteria through chromatin modification of the fungal genome. Since Chávez *et al.* emphasized the high potential of fungi living in harsh environments to produce secondary metabolites with unusual chemical structures<sup>227</sup>, several studies have involved placing microorganisms from extreme environments into contact with one another. Using mixed fermentation, Zhu *et al.* characterized aspergicin from two marine-derived *Aspergillus* strains<sup>228</sup>, and Li *et al.* identified a xanthone derivative<sup>229</sup>. Similarly, the cytotoxic diterpenoids libertellenones A-D were characterized from *Libertella* sp.<sup>230</sup>, and the antibiotic pestalone was identified<sup>231</sup>. Oh *et al.* applied this method to 50 different marine fungi in co-culture with a marine *Actinomycetes* species and characterized emericellamides A and B, with the former displaying low antimicrobial activity and weak cytotoxic activity against a tumor cell line<sup>232</sup>. Regarding other extreme environments, a co-culture of an *A. fumigatus* strain with a bacterium, both isolated from disused mines, led to the characterization of glionitrin A, an antitumor compound with submicromolar cytotoxic activity against four human cancer cell lines<sup>233</sup>.



**Fig. 7** Co-culturing of microorganisms for the identification of secondary metabolites. The signals specifically observed in co-culture (dark-orange image) originate from the secondary metabolites produced only via the communication between the two organisms. The analytical profile of the co-cultures may be obtained by (A) extracting the entire co-culture or (B) specifically analyzing the conflict zone. These profiles are generated by (B1) manual excision of the particular zone or (B2) direct analysis with mass spectrometry (here, Laser Desorption Ionization or Desorption ElectroSpray Ionization).

## 5.2 Modulation of the Physicochemical Growth Conditions

The second environment-dependent strategy consists of abiotic modifications to the culture environment by modifying the light intensity and wavelength, by altering the source of carbon and nitrogen, the ambient pH, the redox status, the culture configuration (static, shaken, liquid or solid) or many other factors<sup>142</sup>. Some studies have highlighted the modulation of fungal secondary metabolism merely according to the quality and source of water. Paranagama *et al.* characterized six fungal secondary metabolites simply by replacing tap water with mineral water in the culture medium<sup>234</sup>. The influence of salinity on secondary metabolomes was also explored. Marine-derived fungi are well known for their ability to produce numerous secondary metabolites<sup>235,236,237</sup>, and particularly antibacterial compounds<sup>238</sup>. An unknown compound with antibacterial activity was found to be produced by *S. elegans*, specifically in culture with 10% salinity<sup>239</sup>. The same experiment with *A. terreus* led to the characterization of three metabolites<sup>240</sup>. Similarly, Zheng *et al.* observed the production of 11 unknown aspochracin-type cyclic tripeptides by growing *A. sclerotiorum* on hypersaline nutrient-limited media<sup>241</sup>.

The oxygen supply is another factor influencing the production of secondary metabolites. Comparing cultures of *Sphaeropsidales* sp. in shaken liquid media (i.e., sufficiently oxygenated cultures) and in static solid media (i.e., weakly oxygenated cultures), Bode *et al.* discovered eight spirobisanthralenes: cladospirones B to I<sup>242</sup>. In another study, the influence of aeration was analyzed, and 120 different cultures of *A. parasiticus* were assessed for antimicrobial activities. The association of bioguided purification and culture condition modulation revealed unknown secondary metabolites with bioactive properties<sup>29</sup>. In another study, the influence of temperature on the metabolome of *A. nidulans* was demonstrated. Among 45 tested conditions involving modulation of the culture temperature, oxygen supply and other parameters, growth at high temperature resulted in the production of two unknown natural compounds with weak antiproliferative effects, aspernidine A and B<sup>243</sup>.

The alteration of carbon and nitrogen sources as well as ~~and~~ the availability of the nutrient source were modulated in several studies. Indeed, Wiemann *et al.* demonstrated that nitrogen availability has a considerable impact on secondary metabolism by affecting the expression of 35 of the 40 gene clusters of *F. fujikuroi*<sup>244</sup>. In some cases, the natural ~~substrate~~ source of carbon and nitrogen allowed the secretion of specific secondary metabolites. For example, *Cephalosporium graminearum* was selected from soil for its antibiotic effect, but its active properties disappeared in a standard medium<sup>245</sup>. Similarly, Overy *et al.* characterized corymbiferone B, a compound only detected in *in planta* cultures of seven *Penicillium* ser. *Corymbifera* sp.<sup>246</sup>. After *in silico* analysis of the *A. nidulans* genome, Scherlack *et al.* noted some genes coding for anthranilate synthases known to synthesize alkaloid precursors. Since alkaloids were never detected from this fungus grown on standard laboratory medium, the fungus was grown on several cereals. Analysis of the extracts focused on aromatic nitrogen-containing metabolites, and four aspoquinolones A-D were characterized from the rice medium<sup>247</sup>. Studying the production of penitrems by *Aspergillus* and *Penicillium* species, Kalinina *et al.* evaluated the impact of different sources of carbon and found that these toxins were mostly produced in media enriched with glucose. The type of nitrogen source also had an impact and glutamate was found to be the most favorable supplement<sup>248</sup>. In contrast, the ~~nutrient~~ source of carbon and nitrogen can be deliberately limited. Sarkar *et al.* cultivated *A. nidulans* under continuous growth with limited supply rates of nitrate, orthophosphate and glucose. Under these conditions, the fungus synthesized a new molecule, preshamixanthone<sup>249</sup>. Similarly, sanghaspirodins A and B were discovered from *A. nidulans* grown with restricted nitrogen resources<sup>250</sup>. To address to the ecologically question related to determining the secondary metabolites produced by a fungal wheat pathogen, our laboratory used a protocol (presented in section 2.2, see reference<sup>41</sup>) that combined a double isotope labeling and the OSMAC principle<sup>41</sup>.

## 6. Conclusions

As described in this review, both targeted genome-based approaches and untargeted strategies are used to characterize fungal secondary metabolites. The different strategies discussed in this review are schematized in Figure 1 and the resulting discoveries of fungal secondary metabolites are listed in Table 1. Genome mining strategies are the most commonly employed strategies for the discovery of new secondary metabolites as they allow investigating the entire metabolic capacity of a fungus, even for gene clusters not expressed in standard laboratory culture conditions. Both the development of new technologies, such as these based on the CRISPR-Cas9 system, and the discovery of new

actors involved in the transcriptional modulation of fungal secondary metabolomes will increase the efficiency, the simplicity and the rapidity of genome-based approaches. The improvement of genome mining strategies needs to gain more insights into the mechanisms involved in the regulation of the biosynthesis of fungal products on the one hand, and to enhance the performance of existing bioinformatics tools for the detection of gene clusters on the other hand.

As highlighted by the One Strain MAAny Compounds principle, the observation of gene clusters enables the assessment of the entire potential of a fungus to generate secondary metabolites. Studies seeking to characterize the secondary metabolites produced by a fungus in particular conditions of growth (i.e., in the context of contact between the fungus and another organism) should consider the growth conditions and focus on the direct analysis of the secondary metabolites. The evolution of analytical instrumentation with higher resolution and sensitivity facilitates the development of untargeted approaches, particularly in the field of natural products research in which the molecules are produced in low quantities and as complex mixtures<sup>251,252</sup>. Aside from the technological advances, several methods, such as the SIL of fungal substrates, were elaborated to specifically target the secondary metabolites and simplify the data mining of complex datasets. New challenges associated with these untargeted metabolomics strategies lie in the development of computational tools for the mining and interpretation of data<sup>199</sup>.

Finally, the evolution of natural product research is moving toward an integrative approach that takes advantage of each tool: bioinformatics for elucidation of the fungal genomic potential, genome engineering for its targeted manipulation and metabolomics for analyzing the global composition of complex samples with high resolution and high sensitivity. These complementary processes were used together in a recent study of *Penicillium* sp. and led to the discovery of previously undescribed compounds<sup>253</sup>. Moreover, an extensive study combining genome mining, chromatin modifications, transcriptomics, proteomics, and metabolomics that also took into account the fungal culture conditions characterized two novel secondary metabolites as virulent factors influencing the destruction of rice fields by *F. fujikuroi*<sup>244</sup>. However, the development of multiomic strategies produces big data and raises the problem of data integration. The continued improvement of the data mining to handle the huge quantities of data produced in this research and the ability to share data on unidentified compounds with other researchers are crucial pursuits. Such integrative studies also necessitate the development of bioinformatics tools to process together all different omics data. Finally, efforts must also be focused on the accessibility and comparability of fungal omics datasets as well as the standardization of the analytical quality<sup>199</sup>.

Strategy (bold) and additional method(s)	Metabolite(s) discovered	Fungal species	References
<b>TARGETED PURIFICATION</b>			
<b>Purification of crude extract</b>			
	Arugosins H and G	<i>Emericella nidulans</i>	17
	Trichodermamides A and B	<i>Trichoderma virens</i>	18
	Isoharzianic acid	<i>Trichoderma harzianum</i>	19
+ Computational analysis with MeHaloCoA	Griseophenone I Chlorogriseofulvin	<i>Penicillium canescens</i>	20
<b>Bioguided purification</b>			
	Penicillin	<i>Penicillium notatum</i>	3
	Fumonisin B1 and B2	<i>Fusarium moniliforme</i>	22
	Besudon	<i>Mollisia benesuada</i>	25
	Trichodermamide A Espergillazine A	<i>Spicaria elegans</i>	23
	Cytochalasins Z <sub>10</sub> -Z <sub>15</sub>	<i>Spicaria elegans</i>	24
	Sorokiniol	<i>Bipolaris sorokiniana</i>	26
	Ligerin	<i>Penicillium</i> strain from sect. <i>Canescentia</i>	27
	Pannomycin	<i>Geomyces pannorum</i>	28
<b>Untargeted metabolomics</b>			
+ Stable Isotope Labeling	9 new compounds	<i>Fusarium graminearum</i>	254
+ Stable Isotope Labeling	A new fumigaclavine analogue	<i>Aspergillus fumigatus</i>	41
+ Stable Isotope Labeling + Computational analysis with molecular networks	4 analogues of nidulanins and fungisporins	<i>Aspergillus nidulans</i>	35
+ Stable Isotope Labeling	82 new secondary metabolites	<i>Penicillium verrucosum</i>	42

+ Adaptation of the growth environment + Computational analysis with molecular networks			
<b>GENOME MINING</b>			
<b>Gene knock-out</b>			
	Emericellamides C-F	<i>Aspergillus nidulans</i>	127
	Aspercryptin	<i>Aspergillus nidulans</i>	132
<b>Heterologous host transfer</b>			
	Neosartoricin B	<i>Trichophyton tonsurans</i>	139
	Terezine D	<i>Aspergillus terreus</i>	144
+ Gene knock-out + Specific pathway overexpression	Asperfuranone	<i>Aspergillus terreus</i>	136
+ Stable isotope labeling	Aculins A and B Epi-aculin A	<i>Aspergillus aculeatus</i>	140
<b>Specific pathway overexpression</b>			
	Aspyridones A and B	<i>Aspergillus nidulans</i>	152
	Fusarielins, F, G and H	<i>Fusarium graminearum</i>	153
	Neosartoricin	<i>Aspergillus fumigatus</i> and <i>Neosartorya fischeri</i>	154
	Apicidin F	<i>Fusarium fujikuroi</i>	155
	7 new metabolites derived from the $\beta$ -resorcyraldehyde	<i>Aspergillus nidulans</i>	148
	Microperfuranone	<i>Aspergillus nidulans</i>	158
	Fellutamide B	<i>Aspergillus nidulans</i>	157
+ Gene knock-out	Asperfuranone	<i>Aspergillus nidulans</i>	133
<b>Global transcription modulation</b>			
	Terrequinone A	<i>Aspergillus nidulans</i>	161
<b>Mutagenesis strategies</b>			
	Purpurogemutant Purpurogemutantidin	<i>Penicillium purpurogenum</i>	30
	Penicimutanolone Penicimutanin A and B Penicimutatin	<i>Penicillium purpurogenum</i>	175
	Antineocyclocitrinols A and B 23-O-methylantineocyclocitrinol	<i>Penicillium purpurogenum</i>	176
	Epiremisorpine B and B1 Isoconiochaetone C	<i>Penicillium purpurogenum</i>	177
	Janthinone Fructigenine A Aspterric acid methyl ester	<i>Penicillium purpurogenum</i>	172
	Curvularin Penicitrinone A Erythro-23-O-methylneocyclocitrinol 22E-7 $\alpha$ -methoxy-5 $\alpha$ ,6 $\alpha$ -epoxyergosta-8(14),22-dien-3 $\beta$ -ol	<i>Penicillium purpurogenum</i>	173
	Penicimutide Cyclo(L-Val-L-Pro) Cyclo(L-Ile-L-Pro) Cyclo(L-Leu-L-Pro) Cyclo(L-Phe-L-Pro)	<i>Penicillium purpurogenum</i>	174
	Phenethyl 5-oxo-L-prolinate Cyclo(D-Pro-D-Phe) Cyclo(D-Tyr-D-Pro) 3 $\beta$ ,5 $\alpha$ ,9 $\alpha$ -trihydroxy-(22E,24R)-ergosta-7,22-dien-6-one	<i>Aspergillus versicolor</i>	178
	Mutolide	<i>Sphaeropsidales</i> sp.	167
<b>Chromatin modification</b>			
	(7S)-(+)-7-O-methylsydonol (7S,11S)-(+)-12-hydroxysydonic acid 7-deoxy-7,14-didehydroxydonol (S)-(+)-sydonol	<i>Aspergillus sydowii</i>	184
	Nigeron A	<i>Aspergillus niger</i>	191
	3 oxylipins Cladochromes F and G	<i>Cladosporium cladosporioides</i>	185
	Lunalides A and B	<i>Diatrype</i> sp.	185
	Atlantinones A and B	<i>Penicillium citreonigrum</i>	186

+ Heterologous host transfer + Transcriptomics	(1R,4R,5S)-guaia-6,10(14)-diene	<i>Fusarium fujikuroi</i>	188
+ Untargeted metabolomics	Antibiotic 1656-G Antibiotic 3127	<i>Aspergillus nidulans</i>	192
<b>TRANSCRIPTOMICS</b>			
	Nidulanin A	<i>Aspergillus nidulans</i>	202
<b>MODIFICATION OF GROWTH ENVIRONMENT</b>			
<b>Interspecies cross-talk</b>			
+ Bioguided purification	Pestalone	<i>Pestalotia</i> (CNL-365) (+ bacterial strain CNJ-328)	231
+ Bioguided purification	4 Mycoalexins	<i>Eutypa lata</i> (+ <i>Botryosphaeria obtuse</i> )	78
+ Purification of crude extract	Secopenicillide C	<i>Penicillium pinophilum</i> (+ <i>Trichoderma harzianum</i> )	221
+ Purification of crude extract	Subenniatins A and B	<i>Fusarium tricinctum</i> (+ <i>Fusarium begonia</i> )	222
+ Purification of crude extract	Macrocapon C And 1 derived of anthranilic acid	<i>Fusarium tricinctum</i> (+ <i>Bacillus subtilis</i> )	225
+ Purification of crude extract	4"-hydroxysulfoxy-2,2"-dimethylthiavin P Its nonsulfated form And 3 other related compounds	<i>Trichophyton rubrum</i> (+ <i>Bionectria ochroleuca</i> )	255
+ Purification of crude extract	Fumiformamide N,N <sub>0</sub> -((1Z,3Z)-1,4-bis(4-methoxyphenyl)buta- 1,3-diene-2,3-diyl)di-formamide	<i>Aspergillus fumigatus</i> (+ <i>Streptomyces peucetius</i> )	223
+ Purification of crude extract	Aspergicin	<i>Aspergillus</i> sp. (+ another <i>Aspergillus</i> sp.)	228
+ Purification of crude extract	8-hydroxy-3-methyl-9-oxo-9H-xanthene-1- carboxylic acid methyl ether	Strain No. K38 (+ strain No. E33)	229
+ Purification of crude extract	Libertellenone A-D	<i>Libertella</i> sp. (+ a marine $\alpha$ -proteobacterium)	230
	Acremostatins A, B and C	<i>Acremonium</i> sp. (+ <i>Mycogone rosea</i> )	224
+ Purification of crude extract	Emericellamides A and B	<i>Emericella</i> sp. (+ <i>Salinispora arenicola</i> )	232
+ Purification of crude extract	Glionitrin A	<i>Aspergillus fumigatus</i> (+ a <i>Sphingomonas</i> bacterial strain)	233
+ Purification of crude extract + Genome mining + Transcriptomics	Fumicycline A and B	<i>Aspergillus fumigatus</i> (+ <i>Streptomyces rapamycinus</i> )	226
+ Computational analysis with molecular networks	2 dimeric phenazines And other bioconversions of bacterial compounds by the fungus	<i>Aspergillus fumigatus</i> (+ <i>Pseudomonas aeruginosa</i> )	256
<b>Modification of physicochemical growth conditions</b>			
+ Bioguided purification	New antibiotics	<i>Aspergillus parasiticus</i>	29
+ Purification of crude extract	Cytosporones F-I Quadrisepin A 5'-hydroxymonocillin III	<i>Paraphaeosphaeria quadrisepinata</i>	234
+ Purification of crude extract	(2E,2'Z)-3,3'-(6,6'-dihydroxybiphenyl-3,3'- diyl)diacrylic acid	<i>Spicaria elegans</i>	239
+ Purification of crude extract	Terremides A and B Terrelactone A	<i>Aspergillus terreus</i>	240
+ Purification of crude extract	Sclerotiotides A-K	<i>Aspergillus sclerotiorum</i>	241
+ Purification of crude extract	Cladospirones B-I	<i>Sphaeropsidales</i> sp.	242
+ Purification of crude extract	Aspernidines A and B	<i>Aspergillus nidulans</i>	243
+ Purification of crude extract	Corymbiferone B	7 <i>Penicillium</i> ser. <i>Corymbifera</i>	246
+ Purification of crude extract	Aspoquinolones A-D	<i>Aspergillus nidulans</i>	247
+ Purification of crude extract + Transcriptomics	Pre-shamixanthone	<i>Aspergillus nidulans</i>	249
+ Purification of crude extract + Transcriptomics + Genome mining	Sanghaspirodins A and B	<i>Aspergillus nidulans</i>	250
+ Transcriptomics + Genome mining + Proteomics + Phylogenetics	4 new metabolites	<i>Fusarium fujikuroi</i>	244

**Table 1.** Strategies used for fungal secondary metabolite discovery.



## 8. Acknowledgements

The platform MetaToul-AXIOM is integrated in the French national infrastructure of metabolomics and fluxomics: MetaboHUB (MetaboHUB-ANR-11-INBS-0010). The authors are grateful to Dr. Rougeulle from the University Paris Diderot (CEDC/CNRS, Paris, France) for her kind help and her expertise in genetics and epigenetics.

## 9. Funding sources

There are no conflicts of interest to declare. This study was co-funded by INRA and French Minister of Higher Education and Research in the context of a project supported by French National Agency of Research (ANR-15-CE21-0010-21Newmyco). TH was supported by a Fellowship from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie.

## 10. References

- 1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2006, **70**, 461–477.
- 2 A. L. Demain and A. Fang, *The natural functions of secondary metabolites*, Springer B., 2000.
- 3 A. Fleming, *Br. J. Exp. Pathol.*, 1929, **10**, 226–236.
- 4 D. W. Denning and M. J. Bromley, *Science (80-. )*, 2015, **347**, 1414–1416.
- 5 G. C. Merhoff and J. M. Porter, *Ann. Surg.*, 1974, **180**, 773.
- 6 N. Bradburn, R. D. Coker and G. Blunden, *Phytochemistry*, 1994, **35**, 817.
- 7 S. Brase, A. Encinas, J. Keck and C. F. Nising, *Chem. Rev.*, 2009, **109**, 3903–3990.
- 8 J. W. Bennett and M. Klich, *Clin. Microbiol. Rev.*, 2003, **16**, 497–516.
- 9 M. C. Fisher, D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw and S. J. Gurr, *Nature*, 2013, **484**, 1–18.
- 10 M. Blackwell, *Am. J. Bot.*, 2011, **98**, 426–438.
- 11 G. F. Bills and J. B. Gloer, *Microbiol Spectr*, 2016, 4, Article Number: UNSP FUNK-0009-2016
- 12 E. K. Prince and G. Pohnert, *Anal Bioanal Chem*, 2010, **396**, 193–197.
- 13 H. V. K. Wangun, H.-M. Dahse and C. Hertweck, *J. Nat. Prod.*, 2007, **70**, 1800–1803.
- 14 W.-L. Wang, Z.-Y. Lu, H.-W. Tao, T.-J. Zhu, Y.-C. Fang, Q.-Q. Gu and W.-M. Zhu, *J. Nat. Prod.*, 2007, **70**, 1558–1564.
- 15 A. W. Alberts, J. Chen, G. Kuron, V. Hunt, J. Huff, C. Hoffman, J. Rothrock, M. Lopez, H. Joshua, E. Harris, A. Patchett, R. Monaghan, S. Currie, E. Stapley, G. Albers-Schonberg, O. Hensens, J. Hirshfield, K. Hoogsteen, J. Liesch and J. Springer, *PNAS*, 1980, **77**, 3957–3961.
- 16 L. Ding, H. M. Dahse and C. Hertweck, *J. Nat. Prod.*, 2012, **75**, 617–621.
- 17 A. Kralj, S. Kehraus, A. Krick, E. Eguereva, G. Kelter, M. Maurer, A. Wortmann, H. H. Fiebig and G. M. König, *J. Nat. Prod.*, 2006, **69**, 995–1000.
- 18 E. Garo, C. M. Starks, P. R. Jensen, W. Fenical, E. Lobkovsky and J. Clardy, *J. Nat. Prod.*, 2003, **66**, 423–426.
- 19 F. Vinale, G. Manganiello, M. Nigro, P. Mazzei, A. Piccolo, A. Pascale, M. Ruocco, R. Marra, N. Lombardi, S. Lanzuise, R. Varlese, P. Cavallo, M. Lorito and S. L. Woo, *Molecules*, 2014, **19**, 9760–9772.
- 20 C. Roullier, Y. Guitton, M. Valery, S. Amand, S. Prado, T. Robiou du Pont, O. Grovel and Y. F. Pouchus, *Anal. Chem.*, 2016, **88**, 9143–9150.
- 21 G. W. Gribble, *Mar. Drugs*, 2015, **13**, 4044–4136.
- 22 W. C. A. Gelderblom, K. Jaskiewicz, W. F. O. Marasas, P. G. Thiel, R. M. Horak, R. Vleggaar and N. P. J. Kriek, *Appl. Environ. Microbiol.*, 1988, **54**, 1806–1811.
- 23 R. Liu, Q.-Q. Gu, W.-M. Zhu, C.-B. Cui and G.-T. Fan, *Arch Pharm Res*, 2005, **28**, 1042–1046.
- 24 R. Liu, Z. Lin, T. Zhu, Y. Fang, Q. Gu and W. Zhu, *J. Nat. Prod.*, 2008, **71**, 1127–1132.
- 25 E. Thines, W.-R. Arendholz, H. Anke and O. Sterner, *J. Antibiot. (Tokyo)*, 1997, **50**, 13–17.
- 26 L. Ali, A. L. Khan, J. Hussain, A. Al-Harrasi, M. Waqas, S.-M. Kang, A. Al-Rawahi and I.-J. Lee, *BMC Microbiol.*, 2016, **16**, 1.
- 27 M. Vansteelandt, E. Blanchet, M. Egorov, F. Petit, L. Toupet, A. Bondon, F. Monteau, B. Bizec, O. P. Thomas, Y. F. Pouchus, R. Le Bot and O. Grovel, *J. Nat. Prod.*, 2013, **76**, 297–301.
- 28 C. A. Parish, M. De La Cruz, S. K. Smith, D. Zink, J. Baxter, S. Tucker-Samaras, J. Collado, G. Platas, G. Bills, M. T. Diez, F. Vicente, F. Pelaez and K. Wilson, *J. Nat. Prod.*, 2009, **72**, 59–62.
- 29 A. A. P. Bracarense and J. A. Takahashi, *Brazilian J. Microbiol.*, 2014, **45**, 313–321.
- 30 S. M. Fang, C. Bin Cui, C. W. Li, C. J. Wu, Z. J. Zhang, L. Li, X. J. Huang and W. C. Ye, *Mar. Drugs*, 2012, **10**, 1266–1287.

- 31 J. Smedsgaard and J. Nielsen, *J. Exp. Bot.*, 2005, **56**, 273–286.
- 32 S. Neumann and S. Böcker, *Anal. Bioanal. Chem.*, 2010, **398**, 2779–2788.
- 33 T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23–60.
- 34 J.-L. Wolfender, G. Marti, A. Thomas and S. Bertrand, *J. Chromatogr. A*, 2015, **1382**, 136–164.
- 35 A. Klitgaard, J. B. Nielsen, R. J. N. Frandsen, M. R. Andersen and K. F. Nielsen, *Anal. Chem.*, 2015, **87**, 6520–6526.
- 36 C. Bueschl, B. Kluger, M. Lemmens, G. Adam, G. Wiesenberger, V. Maschietto, A. Marocco, J. Strauss, S. Bödi, G. G. Thallinger, R. Krska and R. Schuhmacher, *Metabolomics*, 2014, **10**, 754–769.
- 37 C. Bueschl, R. Krska, B. Kluger and R. Schuhmacher, *Anal. Bioanal. Chem.*, 2013, **405**, 27–33.
- 38 F. A. de Jong and C. Beecher, *Bioanalysis*, 2012, **4**, 2303–2314.
- 39 C. S. Clendinen, G. S. Stupp, R. Ajredini, B. Lee-McMullen, C. Beecher and A. S. Edison, *Front. Plant Sci.*, 2015, **6**, 611.
- 40 X. Huang, Y.-J. Chen, K. Cho, I. Nikolskiy, P. A. Crawford and G. J. Patti, *Anal. Chem.*, 2014, **86**, 1632–1639.
- 41 P. M. Cano, E. L. Jamin, S. Tadrist, P. Bourdaud’Hui, M. Péan, L. Debrauwer, I. P. Oswald, M. Delaforge and O. Puel, *Anal. Chem.*, 2013, **85**, 8412–8420.
- 42 T. Hautbergue, O. Puel, S. Tadrist, L. Meneguetti, M. Péan, M. Delaforge, L. Debrauwer, I. P. Oswald and E. L. Jamin, *J. Chrom. B : Biomed. Sci. Appl.*, 2017, **1071**, 29–43
- 43 M. T. Henke and N. L. Kelleher, *Nat. Prod. Rep.*, 2016, **33**, 942–950.
- 44 K. F. Nielsen, T. O. Larsen, N. Keller and T. S. Bugni, *Front. Microbiol.*, 2015, **6**, 71.
- 45 S. P. Gaudêncio and F. Pereira, *Nat. Prod. Rep.*, 2015, **32**, 779–810.
- 46 H. Laatsch, Antibase 2012: The natural compound identifier, <https://application.wiley-vch.de/stmdata/antibase.php>, (accessed 2 January 2017).
- 47 Chapman, 2005.
- 48 T. El-Elimat, M. Figueroa, B. M. Ehrmann, N. B. Cech, C. J. Pearce and N. H. Oberlies, *J. Nat. Prod.*, 2013, **76**, 1709–1716.
- 49 K. F. Nielsen and J. Smedsgaard, *J. Chromatogr. A*, 2003, **1002**, 111–136.
- 50 S. Kildgaard, M. Mansson, I. Dosen, A. Klitgaard, J. C. Frisvad, T. O. Larsen and K. F. Nielsen, *Mar. Drugs*, 2014, **12**, 3681–3705.
- 51 H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703–714.
- 52 Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M. Y. Hirai and K. Saito, *Phytochemistry*, 2012, **82**, 38–45.
- 53 The National Institute of Standards and Technology. NIST standard reference database 1A, <http://www.nist.gov/srd/nist1a.cfm>, (accessed 1 October 2016).
- 54 J. Y. Yang, L. M. Sanchez, C. M. Rath, X. Liu, P. D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. De Felicio, A. Fenner, W. R. Wong, R. G. Linington, L. Zhang, H. M. Debonsi, W. H. Gerwick and P. C. Dorrestein, *J. Nat. Prod.*, 2013, **76**, 1686–1699.
- 55 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrew, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O’Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C.

- Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.
- 56 H. Mohimani, A. Gurevich, A. Mikheenko, N. Garg, L.-F. Nothias, A. Ninomiya, K. Takada, P. C. Dorrestein and P. A. Pevzner, *Nat. Chem. Biol.*, 2016, **13**, 30–37.
- 57 J. Ng, N. Bandeira, W. Liu, M. Ghassemian, T. L. Simmons, W. Gerwick, R. Linington, P. Dorrestein and P. Pevzner, *Nat Methods*, 2009, **6**, 596–599.
- 58 A. Ibrahim, L. Yang, C. Johnston, X. Liu, B. Ma and N. a Magarvey, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 19196–201.
- 59 C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 779–787.
- 60 W. Demuth, M. Karlovits and K. Varmuza, *Anal. Chim. Acta*, 2004, **516**, 75–85.
- 61 F. Olivon, G. Grelier, F. Roussi, M. Litaudon and D. Touboul, *Anal. Chem.*, 2017, **89**, 7836–7840.
- 62 H. Mohimani and P. A. Pevzner, *Nat. Prod. Rep.*, 2016, **33**, 73–86.
- 63 J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V Burgess and S. Rogers, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 13738–13743.
- 64 S. Böcker and K. Dührkop, *J. Cheminform.*, 2016, **8**, 1–26.
- 65 A. Vaniya and O. Fiehn, *TrAC Trends Anal. Chem.*, 2015, **69**, 52–61.
- 66 F. Rasche, A. Svatos, R. K. Maddula, C. Boettcher and S. Boecker, *Anal. Chem.*, 2011, **83**, 1243–1251.
- 67 F. Rasche, K. Scheubert, F. Hufsky, T. Zichner, M. Kai, A. Svatos and S. Bocker, *Anal. Chem.*, 2012, **84**, 3417–3426.
- 68 H. Mohimani, W.-T. Liu, R. D. Kersten, B. S. Moore, P. C. Dorrestein and P. A. Pevzner, *J. Nat. Prod.*, 2014, **77**, 1902–1909.
- 69 T. H. J. Niedermeyer and M. Strohal, *PLoS One*, 2012, **7**, e44913.
- 70 K. F. Medzihradszky and R. J. Chalkley, *Mass Spectrom. Rev.*, 2015, **34**, 43–63.
- 71 K. Suhre and P. Schmitt-Kopplin, *Nucleic Acids Res.*, 2008, **36**, 481–484.
- 72 K. Dührkop, H. Shen, M. Meusel, J. Rousu and S. Böcker, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12580–5.
- 73 F. Allen, A. Pon, M. Wilson, R. Greiner and D. Wishart, *Nucleic Acids Res.*, 2014, **42**, 1–6.
- 74 F. Allen, R. Greiner and D. Wishart, *Metabolomics*, 2015, **11**, 98–110.
- 75 L. Ridder, J. van der Hooft, S. Verhoevem, R. C. H. de Vos, R. J. Bino and J. Vervoort, *Anal. Chem.*, 2013, **85**, 6033–6040.
- 76 S. Wolf, S. Schmidt, M. Müller-Hannemann and S. Neumann, *BMC Bioinformatics*, 2010, **11**, 148.
- 77 S. Bertrand, O. Schumpp, N. Bohni, A. Bujard, A. Azzollini, M. Monod, K. Gindro and J. L. Wolfender, *J. Chromatogr. A*, 2013, **1292**, 219–228.
- 78 G. Glauser, K. Gindro, J. Fringeli, J. P. De Joffrey, S. Rudaz and J.-L. Wolfender, *J. Agric. Food Chem.*, 2009, **57**, 1127–1134.
- 79 B. C. Covington, J. A. McLean and B. O. Bachmann, *Nat. Prod. Rep.*, 2016, **34**, 6–24.
- 80 H. Oberacher, M. Pavlic, K. Libiseller, B. Schubert, M. Sulyok, R. Schuhmacher, E. Csaszar and H. C. Köfeler, *J. Mass Spectrom.*, 2009, **44**, 494–502.
- 81 M. Machida, K. Asai, M. Sano, T. Tanaka, T. Kumagai, G. Terai, K. Kusumoto, T. Arima, O. Akita, Y. Kashiwagi, K. Abe, K. Gomi, H. Horiuchi, K. Kitamoto, T. Kobayashi, M. Takeuchi, D. W. Denning, J. E. Galagan, W. C. Nierman, J. Yu, D. B. Archer, J. W. Bennett, D. Bhatnagar, T. E. Cleveland, N. D. Fedorova, O. Gotoh, H. Horikawa, A. Hosoyama, M. Ichinomiya, R. Igarashi, K. Iwashita, P. R. Juvvadi, M. Kato, Y. Kato, T. Kin, A. Kokubun, H. Maeda, N. Maeyama, J. Maruyama, H. Nagasaki, T. Nakajima, K. Oda, K. Okada, I. Paulsen, K. Sakamoto, T. Sawano, M. Takahashi, K. Takase, Y. Terabayashi, J. R. Wortman, O. Yamada, Y. Yamagata, H. Anazawa, Y. Hata, Y. Koide, T. Komori, Y. Koyama, T. Minetoki, S. Suharnan, A. Tanaka, K. Isono, S. Kuhara, N. Ogasawara and H. Kikuchi, *Nature*, 2005, **438**, 1157–1161.
- 82 C. L. Schardl, *Chem. Biol.*, 2006, **13**, 5–6.
- 83 P. Wiemann and N. P. Keller, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 301–313.
- 84 A. A. Brakhage and V. Schroeckh, *Fungal Genet. Biol.*, 2011, **48**, 15–22.
- 85 H. Gross, *Appl. Microbiol. Biotechnol.*, 2007, **75**, 267–277.
- 86 M. Zerikly and G. L. Challis, *ChemBioChem*, 2009, **10**, 625–633.
- 87 J. M. Winter, S. Behnken and C. Hertweck, *Curr. Opin. Chem. Biol.*, 2011, **15**, 22–31.
- 88 K. Ochi, *J. Antibiot. (Tokyo)*, 2016, 1–16.
- 89 N. Ziemert, M. Alanjary and T. Weber, *Nat. Prod. Rep.*, 2016, **33**, 988–1005.
- 90 K. Ochi and T. Hosaka, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 87–98.

- 91 T. Weber and H. U. Kim, *Synth. Syst. Biotechnol.*, 2016, **1**, 69–79.
- 92 T. A. J. Van der Lee and M. H. Medema, *Fungal Genet. Biol.*, 2016, **89**, 29–36.
- 93 A. K. Chavali and S. Y. Rhee, *Brief. Bioinform.*, 2017, 1–13.
- 94 N. Khaldi, F. T. Seifuddin, G. Turner, D. Haft, W. C. Nierman, K. H. Wolfe and N. D. Fedorova, *Fungal Genet. Biol.*, 2010, **47**, 736–741.
- 95 A. Starcevic, J. Zucko, J. Simunkovic, P. F. Long, J. Cullum and D. Hranueli, *Nucleic Acids Res.*, 2008, **36**, 6882–6892.
- 96 M. H. Medema, K. Blin, P. Cimermancic, V. De Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–346.
- 97 K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204–W2012.
- 98 T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Brucocoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Res.*, 2015, **43**, W237–W243.
- 99 M. A. Skinnider, C. A. Dejong, P. N. Rees, C. W. Johnston, H. Li, A. L. H. Webster, M. A. Wyatt and N. A. Magarvey, *Nucleic Acids Res.*, 2015, **43**, 9645–9662.
- 100 M. H. Medema, E. Takano and R. Breitling, *Mol. Biol. Evol.*, 2013, **30**, 1218–1223.
- 101 T. Wolf, V. Shelest, N. Nath and E. Shelest, *Bioinformatics*, 2016, **32**, 1138–1143.
- 102 I. Takeda, M. Umemura, H. Koike, K. Asai and M. Machida, *DNA Res.*, 2014, **21**, 447–457.
- 103 J. Tannous, R. El Khoury, S. P. Snini, Y. Lippi, A. El Khoury, A. Atoui, R. Lteif, I. P. Oswald and O. Puel, *Int. J. Food Microbiol.*, 2014, **189**, 51–60.
- 104 M. Umemura, H. Koike, N. Nagano, T. Ishii, J. Kawano, N. Yamane, I. Kozone, K. Horimoto, K. Shin-ya, K. Asai, J. Yu, J. W. Bennett and M. Machida, *PLoS One*, 2013, **8**, e84028.
- 105 M. Umemura, H. Koike and M. Machida, *Front. Microbiol.*, 2015, **6**, 371.
- 106 P. Wiemann, C.-J. Guo, J. M. Palmer, R. Sekonyela, C. C. C. Wang and N. P. Keller, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 17065–70.
- 107 K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. De Los Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2017, **45**, W36–W41.
- 108 T. J. Sharpton, J. E. Stajich, S. D. Rounsley, M. J. Gardner, J. R. Wortman, V. S. Jordar, R. Maiti, C. D. Kodira, D. E. Neafsey, Q. Zeng, C.-H. Hung, C. McMahan, A. Muszewska, M. Grynberg, M. A. Mandel, E. M. Kellner, B. M. Barker, J. N. Galagiani, M. J. Orbach, T. N. Kirkland, G. T. Cole, M. R. Henn, B. W. Birren and J. W. Taylor, *Genome Res.*, 2009, **19**, 1722.
- 109 D. W. Brown and R. H. Proctor, *Fungal Genet. Biol.*, 2016, **89**, 37–51.
- 110 K. Throckmorton, P. Wiemann and N. P. Keller, *Toxins (Basel)*, 2015, **7**, 3572–3607.
- 111 R. A. Syme, J. K. Hane, T. L. Friesen and R. P. Oliver, *G3 Genes/ Genomes/ Genet.*, 2013, **3**, 959–969.
- 112 B. J. Condon, Y. Leng, D. Wu, K. E. Bushley, R. A. Ohm, R. Otilar, J. Martin, W. Schackwitz, J. Grimwood, N. A. I. MohdZainudin, C. Xue, R. Wang, V. A. Manning, B. Dhillon, Z. J. Tu, B. J. Steffenson, A. Salamov, H. Sun, S. Lowry, K. LaButti, J. Han, A. Copeland, E. Lindquist, K. Barry, J. Schmutz, S. E. Baker, L. M. Ciuffetti, I. V. Grigoriev, S. Zhong and B. G. Turgeon, *PLoS Genet.*, 2013, **9**, e1003233.
- 113 Y. H. Chooi and Y. Tang, *J. Org. Chem.*, 2012, **77**, 9933–9953.
- 114 H. Tae, E.-B. Kong and K. Park, *BMC Bioinformatics*, 2007, **8**, 327.
- 115 M. H. Li, P. M. Ung, J. Zajkowski, S. Garneau-Tsodikova and D. H. Sherman, *BMC Bioinformatics*, 2009, **10**, 1.
- 116 S. Anand and D. Mohanty, in *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications*, IGI Global snippet, USA: IGI G., 2011, pp. 414–439.
- 117 G. Yadav, R. S. Gokhale and D. Mohanty, *Nucleic Acids Res.*, 2003, **31**, 3654–3658.
- 118 M. Z. Ansari, G. Yadav, R. S. Gokhale and D. Mohanty, *Nucleic Acids Res.*, 2004, **32**, W405–W413.
- 119 P. Kamra, R. S. Gokhale and D. Mohanty, *Nucleic Acids Res.*, 2005, **33**, W220–W225.
- 120 B. O. Bachmann and J. Ravel, *Methods Enzymol.*, 2009, **458**, 181–217.
- 121 S. Anand, M. V. R. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari and D. Mohanty, *Nucleic Acids Res.*, 2010, **38**, W487–W496.
- 122 M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, 362–367.
- 123 C. Prieto, *Characterization of nonribosomal peptide synthetases with NRPSsp*, Springer P., 2016.
- 124 C. Prieto, C. Garcia-estrada, D. Lorenzana and J. F. Martin, *Bioinformatics*, 2012, **28**, 426–427.

- 125 Y. Minowa, M. Araki and M. Kanehisa, *J. Mol. Biol.*, 2007, **368**, 1500–1517.
- 126 M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K.-D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linqun, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. A. Neilan, M. Nett, J. Nielsen, F. O’Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süßmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano and F. O. Glöckner, *Nat. Chem. Biol.*, 2015, **11**, 625–631.
- 127 Y. M. Chiang, E. Szewczyk, T. Nayak, A. D. Davidson, J. F. Sanchez, H. C. Lo, W. Y. Ho, H. Simityan, E. Kuo, A. Praseuth, K. Watanabe, B. R. Oakley and C. C. C. Wang, *Chem. Biol.*, 2008, **15**, 527–532.
- 128 Y. Ninomiya, K. Suzuki, C. Ishii and H. Inoue, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12248–12253.
- 129 S. Krappmann, C. Sasse and G. H. Braus, *Eukaryot. Cell*, 2006, **5**, 212–215.
- 130 S. Pöggeler and U. Kück, *Gene*, 2006, **378**, 1–10.
- 131 M. E. Da Silva Ferreira, M. R. V. Z. Kress, M. Savoldi, M. H. S. Goldman, A. Härtl, T. Heinekamp, A. A. Brakhage and G. H. Goldman, *Eukaryot. Cell*, 2006, **5**, 207–211.
- 132 Y. M. Chiang, M. Ahuja, C. E. Oakley, R. Entwistle, A. Asokan, C. Zutz, C. C. C. Wang and B. R. Oakley, *Angew. Chemie - Int. Ed.*, 2016, **55**, 1662–1665.
- 133 Y. M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley and C. C. C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 2965–2970.
- 134 Y. M. Chiang, S. L. Chang, B. R. Oakley and C. C. C. Wang, *Curr. Opin. Chem. Biol.*, 2011, **15**, 137–143.
- 135 L. B. Pickens, Y. Tang and Y.-H. Chooi, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 211.
- 136 Y. M. Chiang, C. E. Oakley, M. Ahuja, R. Entwistle, A. Schultz, S. L. Chang, C. T. Sung, C. C. C. Wang and B. R. Oakley, *J. Am. Chem. Soc.*, 2013, **135**, 7720–7731.
- 137 J. Yaegashi, B. R. Oakley and C. C. C. Wang, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 433–442.
- 138 K. Sakai, H. Kinoshita, T. Shimizu and T. Nihira, *J. Biosci. Bioeng.*, 2008, **106**, 466–472.
- 139 W. B. Yin, Y. H. Chooi, A. R. Smith, R. A. Cacho, Y. Hu, T. C. White and Y. Tang, *ACS Synth. Biol.*, 2013, **2**, 629–634.
- 140 L. M. Petersen, D. K. Holm, C. H. Gotfredsen, U. H. Mortensen and T. O. Larsen, *ChemBioChem*, 2015, **16**, 2200–2204.
- 141 B. A. Pfeifer and C. Khosla, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 106–118.
- 142 K. Scherlach and C. Hertweck, *Org. Biomol. Chem.*, 2009, **7**, 1753–1760.
- 143 G. L. Challis, *Microbiology*, 2008, **154**, 1555–1569.
- 144 J. W. Bok, R. Ye, K. D. Clevenger, D. Mead, M. Wagner, A. Krerowicz, J. C. Albright, A. W. Goering, P. M. Thomas, N. L. Kelleher, N. P. Keller and C. C. Wu, *BMC Genomics*, 2015, **16**, 343.
- 145 K. D. Clevenger, J. W. Bok, R. Ye, G. P. Miley, M. H. Verdán, T. Velk, C. Chen, K. Yang, M. T. Robey, P. Gao, M. Lamprecht, P. M. Thomas, M. N. Islam, J. M. Palmer, C. C. Wu, N. P. Keller and N. L. Kelleher, *Nat. Chem. Biol.*, 2017, **13**, 895–901.
- 146 Y. H. Chooi and P. S. Solomon, *Front. Microbiol.*, 2014, **5**, 640.
- 147 A. A. Brakhage, *Nat. Rev. Microbiol.*, 2013, **11**, 21–32.

- 148 M. Ahuja, Y. M. Chiang, S. L. Chang, M. B. Praseuth, R. Entwistle, J. F. Sanchez, H. C. Lo, H. H. Yeh, B. R. Oakley and C. C. C. Wang, *J. Am. Chem. Soc.*, 2012, **134**, 8212–8221.
- 149 K. Shimizu, J. K. Hicks, T. P. Huang and N. P. Keller, *Genetics*, 2003, **165**, 1095–1104.
- 150 P. K. Chang, K. C. Ehrlich, J. Yu, D. Bhatnagar and T. E. Cleveland, *Appl. Environ. Microbiol.*, 1995, **61**, 2372–2377.
- 151 F. Y. Lim, J. F. Sanchez, C. C. C. Wang and N. P. Keller, *Methods Enzymol.*, 2012, **517**, 303.
- 152 S. Bergmann, J. Schümann, K. Scherlach, C. Lange, A. A. Brakhage and C. Hertweck, *Nat. Chem. Biol.*, 2007, **3**, 213–217.
- 153 J. L. Sorensen, F. T. Hansen, T. E. Sondergaard, D. Staerk, T. V. Lee, R. Wimmer, L. G. Klitgaard, S. Purup, H. Giese and R. J. N. Frandsen, *Environ. Microbiol.*, 2012, **14**, 1159–1170.
- 154 Y. H. Chooi, J. Fang, H. Liu, S. G. Filler, P. Wang and Y. Tang, *Org. Lett.*, 2013, **15**, 780–783.
- 155 K. W. Von Barga, E. M. Niehaus, K. Bergander, R. Brun, B. Tudzynski and H. U. Humpf, *J. Nat. Prod.*, 2013, **76**, 2136–2140.
- 156 S. Bergmann, A. N. Funk, K. Scherlach, V. Schroeckh, E. Shelest, U. Horn, C. Hertweck and A. A. Brakhage, *Appl. Environ. Microbiol.*, 2010, **76**, 8143–8149.
- 157 H. H. Yeh, M. Ahuja, Y. M. Chiang, C. E. Oakley, S. Moore, O. Yoon, H. Hajovsky, J. W. Bok, N. P. Keller, C. C. C. Wang and B. R. Oakley, *ACS Chem. Biol.*, 2016, **11**, 2275–2284.
- 158 H. H. Yeh, Y. M. Chiang, R. Entwistle, M. Ahuja, K. H. Lee, K. S. Bruno, T. K. Wu, B. R. Oakley and C. C. C. Wang, *Appl. Microbiol. Biotechnol.*, 2012, **96**, 739–748.
- 159 J. W. Bok and N. P. Keller, *Eukaryot. Cell*, 2004, **3**, 527–535.
- 160 R. M. Perrin, N. D. Fedorova, W. B. Jin, R. A. Cramer, J. R. Wortman, H. S. Kim, W. C. Nierman and N. P. Keller, *PLoS Pathog.*, 2007, **3**, e50.
- 161 J. W. Bok, D. Hoffmeister, L. A. Maggio-Hall, R. Murillo, J. D. Glasner and N. P. Keller, *Chem. Biol.*, 2006, **13**, 31–37.
- 162 S. Dhingra, A. L. Lind, H. C. Lin, Y. Tang, A. Rokas and A. M. Calvo, *PLoS One*, 2013, **8**, e77147.
- 163 R. M. Duran, J. W. Cary and A. M. Calvo, *Appl. Microbiol. Biotechnol.*, 2007, **73**, 1158–1168.
- 164 C. E. Oakley, M. Ahuja, W.-W. Sun, R. Entwistle, T. Akashi, J. Yaegashi, C.-J. Guo, G. C. Cerqueira, J. Russo Wortman, C. C. C. Wang, Y.-M. Chiang and B. R. Oakley, *Mol. Microbiol.*, 2016, **103**, 347–365.
- 165 A. N. Patananan, J. M. Palmer, G. S. Garvey, N. P. Keller and S. G. Clarke, *J. Biol. Chem.*, 2013, **288**, 14032–14045.
- 166 Y. Reyes-Dominguez, J. W. Bok, H. Berger, E. K. Shwab, A. Basheer, A. Gallmetzer, C. Scazzocchio, N. Keller and J. Strauss, *Mol. Microbiol.*, 2010, **76**, 1376–1386.
- 167 H. B. Bode, M. Walker and A. Zeeck, *European J. Org. Chem.*, 2000, **2000**, 1451–1456.
- 168 D. J. Downes, M. Chonofsky, K. Tan, B. T. Pfannenstiel, S. L. Reck-Peterson and R. B. Todd, *G3 Genes/ Genomes/ Genet.*, 2014, **4**, 2483–2492.
- 169 V. Ramamoorthy, S. Shantappa, S. Dhingra and A. Calvo, *Mol Microbiol*, 2012, **85**, 795–814.
- 170 K. Ochi, *Biosci. Biotechnol. Biochem.*, 2007, **71**, 1373–1386.
- 171 T. Hosaka, M. Ohnishi-Kameyama, H. Muramatsu, K. Murakami, Y. Tsurumi, S. Kodani, M. Yoshida, A. Fujie and K. Ochi, *Nat. Biotechnol.*, 2009, **27**, 462–464.
- 172 Y. J. Chai, C. Bin Cui, C. W. Li, C. J. Wu, C. K. Tian and W. Hua, *Mar. Drugs*, 2012, **10**, 559–582.
- 173 C. J. Wu, L. Yi, C. Bin Cui, C. W. Li, N. Wang and X. Han, *Mar. Drugs*, 2015, **13**, 2465–2487.
- 174 N. Wang, C. Cui and C.-W. Li, *Arch. Pharm. Res.*, 2016, **39**, 762–770.
- 175 S. M. Fang, C. J. Wu, C. W. Li and C. Bin Cui, *Mar. Drugs*, 2014, **12**, 1788–1814.
- 176 M. W. Xia, C. Bin Cui, C. W. Li and C. J. Wu, *Mar. Drugs*, 2014, **12**, 1545–1568.
- 177 M.-W. Xia, C.-B. Cui, C.-W. Li, C.-J. Wu, J.-X. Peng and D.-H. Li, *Mar. Drugs*, 2015, **13**, 5219–5236.
- 178 Y. Dong, C. Bin Cui, C. W. Li, W. Hua, C. J. Wu, T. J. Zhu and Q. Q. Gu, *Mar. Drugs*, 2014, **12**, 4326–4352.
- 179 J. Sekiguchi and G. M. Gaucher, *Appl. Environ. Microbiol.*, 1977, **33**, 147–158.
- 180 M. S. Ramirez and M. E. Tolmasky, *Drug Resist. Updat.*, 2010, **13**, 151–171.
- 181 E. K. Shwab, J. W. Bok, M. Tribus, J. Galehr, S. Graessle and N. P. Keller, *Eukaryotic Cell*, 2007, **6**, 1656–1664.
- 182 H. W. Nützmann, J. Fischer, K. Scherlach, C. Hertweck and A. A. Brakhage, *Appl. Environ. Microbiol.*, 2013, **79**, 6102–6109.
- 183 R. H. Cichewicz, *Nat. Prod. Rep.*, 2010, **27**, 11–22.
- 184 Y. M. Chung, C. K. Wei, D. W. Chuang, M. El-Shazly, C. T. Hsieh, T. Asai, Y. Oshima, T. J. Hsieh, T. L. Hwang, Y. C. Wu and F. R. Chang, *Bioorganic Med. Chem.*, 2013, **21**, 3866–3872.

- 185 R. B. Williams, J. C. Henrikson, A. R. Hoover, A. E. Lee and R. H. Cichewicz, *Org. Biomol. Chem.*, 2008, **6**, 1895–1897.
- 186 X. Wang, J. G. Sena Filho, A. R. Hoover, J. B. King, T. K. Ellis, D. R. Powell and R. H. Cichewicz, *J. Nat. Prod.*, 2010, **73**, 942–948.
- 187 J. W. Bok, Y.-M. Chiang, E. Szewczyk, Y. Reyes-Dominguez, A. D. Davidson, J. F. Sanchez, H.-C. Lo, K. Watanabe, J. Strauss, B. R. Oakley, C. C. C. Wang and N. P. Keller, *Nat. Chem. Biol.*, 2009, **5**, 462–464.
- 188 L. Studt, S. M. R. Rösler, I. Burkhardt, B. Arndt, M. Freitag, H.-U. Humpf, J. S. Dickschat and B. Tudzynski, *Environ. Microbiol.*, 2016, **18**, 4037–4054.
- 189 A. A. Soukup, Y.-M. Chiang, J. W. Bok, Y. Reyes-Dominguez, B. R. Oakley, C. C. C. Wang, J. Strauss and N. P. Keller, *Mol Microbiol*, 2012, **86**, 314–330.
- 190 S. M. Rösler, K. Kramer, I. Finkemeier, H. U. Humpf and B. Tudzynski, *Mol. Microbiol.*, 2016, **102**, 951–974.
- 191 J. C. Henrikson, A. R. Hoover, P. M. Joyner and R. H. Cichewicz, *Org. Biomol. Chem.*, 2009, **7**, 435–438.
- 192 J. C. Albright, M. T. Henke, A. A. Soukup, R. A. McClure, R. J. Thomson, N. P. Keller and N. L. Kellerher, *ACS Chem. Biol.*, 2015, **10**, 1535–1541.
- 193 M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science (80-. )*, 2012, **337**, 816–822.
- 194 J. A. Doudna and E. Charpentier, *Science (80-. )*, 2014, **346**, 1258096–1258096.
- 195 C. S. Nodvig, J. B. Nielsen, M. E. Kogle and U. H. Mortensen, *PLoS One*, 2015, **10**, 1–18.
- 196 C. Zhang, X. Meng, X. Wei and L. Lu, *Fungal Genet. Biol.*, 2016, **86**, 47–57.
- 197 T. Arazoe, K. Miyoshi, T. Yamato, T. Ogawa, S. Ohsato, T. Arie and S. Kuwata, *Biotechnol. Bioeng.*, 2015, **112**, 2543–2549.
- 198 J. Weber, V. Valiante, C. S. Nødvig, D. J. Mattern, R. A. Slotkowski, U. H. Mortensen and A. A. Brakhage, *ACS Synth. Biol.*, 2017, **6**, 62–68.
- 199 V. Meyer, M. R. Andersen, A. A. Brakhage, G. H. Braus, M. X. Caddick, T. C. Cairns, R. P. de Vries, T. Haarmann, K. Hansen, C. Hertz-Fowler, S. Krappmann, U. H. Mortensen, M. A. Peñalva, A. F. J. Ram and R. M. Head, *Fungal Biol. Biotechnol.*, 2016, **3**, 6.
- 200 M. L. Nielsen, T. Isbrandt, K. B. Rasmussen, U. Thrane, J. B. Hoof, T. O. Larsen and U. H. Mortensen, *PLoS One*, 2017, **12**, e0169712.
- 201 A. Vojta, P. Dobrinic, V. Tadic, L. Bockor, P. Korac, B. Julg, M. Klasic and V. Zoldos, *Nucleic Acids Res.*, 2016, **44**, 5615–5628.
- 202 M. R. Andersen, J. B. Nielsen, A. Klitgaard, L. M. Petersen, M. Zachariasen, T. J. Hansen, L. H. Blicher, C. H. Gotfredsen, T. O. Larsen, K. F. Nielsen and U. H. Mortensen, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E99–E107.
- 203 N. P. Keller, G. Turner and J. W. Bennett, *Nat Rev Microbiol*, 2005, **3**, 937–947.
- 204 S. V. S. Ipcho, J. K. Hane, E. A. Antoni, D. Ahren, B. Henrissat, T. L. Friesen, P. S. Solomon and R. P. Oliver, *Mol. Plant Pathol.*, 2012, **13**, 531–545.
- 205 Y.-H. Chooi, P. S. Solomon, J. M. Palmer and A. Trchounian, *Front. Microbiol.*, 2014, **19**, 640.
- 206 Z. Wang, M. Gerstein and M. Snyder, *Nat. Rev. Genet.*, 2009, **10**, 57–63.
- 207 X.-W. Zhang, L.-J. Jia, Y. Zhang, G. Jiang, X. Li, D. Zhang and W.-H. Tang, *Plant Cell*, 2012, **24**, 5159–5176.
- 208 R. A. Owens, S. Hammel, K. J. Sheridan, G. W. Jones and S. Doyle, *PLoS One*, 2014, **9**, e106942.
- 209 S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai and N. L. Kelleher, *Nat. Biotechnol.*, 2009, **27**, 951–956.
- 210 J. L. Meier and M. D. Burkart, *Curr. Opin. Chem. Biol.*, 2011, **15**, 48–56.
- 211 J. L. López, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, 2007, **849**, 190–202.
- 212 S. Wiese, K. A. Reidegeld, H. E. Meyer and B. Warscheid, *Proteomics*, 2007, **7**, 340–350.
- 213 R. D. Taylor, A. Saparno, B. Blackwell, V. Anoop, S. Gleddie, N. A. Tinker and L. J. Harris, *Proteomics*, 2008, **8**, 2256–2265.
- 214 A. A. Soukup, N. P. Keller and P. Wiemann, *Methods Mol Biol.*, 2016, **1401**, 149–160.
- 215 J. L. Meier, S. Niessen, H. S. Hoover, T. L. Foley, B. F. Cravatt and M. D. Burkart, *ACS Chem. Biol.*, 2009, **4**, 948–957.
- 216 J. L. Meier, A. C. Mercer and M. D. Burkart, *J. Am. Chem. Soc.*, 2008, **130**, 5443–5445.
- 217 S. Bertrand, N. Bohni, S. Schnee, O. Schumpp, K. Gindro and J. L. Wolfender, *Biotechnol. Adv.*,

- 2014, 32, 1180–1204.
- 218 W. Yin and N. P. Keller, *J. Microbiol.*, 2011, **49**, 329–339.
- 219 H. Nützmann, Y. Reyes-dominguez, K. Scherlach, V. Schroeckh and F. Horn, *PNAS*, 2011, **108**, 14282–14287.
- 220 A. Marmann, A. H. Aly, W. Lin, B. Wang and P. Proksch, *Mar. Drugs*, 2014, **12**, 1043–1065.
- 221 K. Nonaka, T. Abe, M. Iwatsuki, M. Mori, T. Yamamoto, K. Shiomi, S. Mura and R. Masuma, *J. Antibiot. (Tokyo)*, 2011, **64**, 769–77491.
- 222 J. P. Wang, W. Lin, V. Wray, D. Lai and P. Proksch, *Tetrahedron Lett.*, 2013, **54**, 2492–2496.
- 223 K. M. Zuck, S. Shipley and D. J. Newman, *J. Nat. Prod.*, 2011, **74**, 1653–1657.
- 224 T. Degenkolb, S. Heinze, B. Schlegel, G. Strobel and U. Grafe, *Biosci. Biotechnol. Biochem.*, 2002, **66**, 883–886.
- 225 A. R. B. Ola, D. Thomy, D. Lai, H. Brötz-Oesterhelt and P. Proksch, *J. Nat. Prod.*, 2013, **76**, 2094–2099.
- 226 C. C. König, K. Scherlach, V. Schroeckh, F. Horn, S. Nietzsche, A. A. Brakhage and C. Hertweck, *ChemBioChem*, 2013, **14**, 938–942.
- 227 R. Chávez, F. Fierro, R. O. García-Rico and I. Vaca, *Front. Microbiol.*, 2015, **6**, 903.
- 228 F. Zhu, G. Chen, X. Chen, M. Huang and X. Wan, *Chem.*, 2011, **47**, 767.
- 229 C. Li, J. Zhang, C. Shao, W. Ding, Z. She and Y. Lin, *Chem. Nat. Compd.*, 2011, **47**, 382.
- 230 D. C. Oh, P. R. Jensen, C. A. Kauffman and W. Fenical, *Bioorganic Med. Chem.*, 2005, **13**, 5267–5273.
- 231 M. Cueto, P. R. Jensen, C. Kauffman, W. Fenical, E. Lobkovsky and J. Clardy, *J. Nat. Prod.*, 2001, **64**, 1444–1446.
- 232 D. C. Oh, C. A. Kauffman, P. R. Jensen and W. Fenical, *J. Nat. Prod.*, 2007, **70**, 515–520.
- 233 H. B. Park, H. C. Kwon, C. H. Lee and H. O. Yang, *J. Nat. Prod.*, 2009, **72**, 248–252.
- 234 P. A. Paranagama, E. M. K. Wijeratne and A. A. L. Gunatilaka, *J. Nat. Prod.*, 2007, **70**, 1939–1945.
- 235 M. E. Rateb and R. Ebel, *Nat. Prod. Rep.*, 2011, **28**, 290–344.
- 236 R. C. Bonugli-Santos, M. R. dos S. Vasconcelos, M. R. Z. Passarini, G. A. L. Vieira, V. C. P. Lopes, P. H. Mainardi, J. A. dos Santos, L. de A. Duarte, I. V. R. Otero, A. M. da S. Yoshida, V. A. Feitosa, A. Pessoa and L. D. Sette, *Front. Microbiol.*, 2015, **6**, 269.
- 237 K. Li, Y.-W. Chung-Davidson, U. Bussy and W. Li, *Mar. Drugs*, 2015, **13**, 2694–2713.
- 238 C. Christophersen, O. Crescente, J. C. Frisvad, L. Gram, J. Nielsen, H. Nielsen and L. Rahbaek, *Mycopathologia*, 1999, **143**, 135–138.
- 239 Y. Wang, Z. Lu, K. Sun and W. Zhu, *Mar. Drugs*, 2011, **9**, 535–542.
- 240 Y. Wang, J. Zheng, P. Liu, W. Wang and W. Zhu, *Mar. Drugs*, 2011, **9**, 1368–1378.
- 241 J. Zheng, Z. Xu, Y. Wang, K. Hong, P. Liu and W. Zhu, *J. Nat. Prod.*, 2010, **73**, 1133–1137.
- 242 H. B. Bode, M. Walker and A. Zeeck, *European J. Org. Chem.*, 2000, **2000**, 3185–3193.
- 243 K. Scherlach, J. Schuemann, H.-M. Dahse and C. Hertweck, *J. Antibiot. (Tokyo)*, 2010, **63**, 375.
- 244 P. Wiemann, C. M. K. Sieber, K. W. von Bargen, L. Studt, E. M. Niehaus, J. J. Espino, K. Huß, C. B. Michielse, S. Albermann, D. Wagner, S. V. Bergner, L. R. Connolly, A. Fischer, G. Reuter, K. Kleigrew, T. Bald, B. D. Wingfield, R. Ophir, S. Freeman, M. Hippler, K. M. Smith, D. W. Brown, R. H. Proctor, M. Münsterkötter, M. Freitag, H. U. Humpf, U. Güldener and B. Tudzynski, *PLoS Pathog.*, 2013, **9**, e1003475.
- 245 G. W. Bruehl, R. L. Millar and B. Cunfer, *Can. J. Plant Sci.*, 1969, **49**, 235–246.
- 246 D. P. Overy, J. Smedsgaard, J. C. Frisvad, R. K. Phipps and U. Thrane, *J. Appl. Microbiol.*, 2006, **101**, 1292–1300.
- 247 K. Scherlach and C. Hertweck, *Org. Biomol. Chem.*, 2006, **4**, 3517–3520.
- 248 S. A. Kalinina, A. Jagels, B. Cramer, R. Geisen and H. U. Humpf, *Toxins (Basel)*, 2017, **9**, 1–22.
- 249 A. Sarkar, A. N. Funk, K. Scherlach, F. Horn, V. Schroeckh, P. Chankhamjon, M. Westermann, M. Roth, A. A. Brakhage, C. Hertweck and U. Horn, *J. Biotechnol.*, 2012, **160**, 64–71.
- 250 K. Scherlach, A. Sarkar, V. Schroeckh, H. M. Dahse, M. Roth, A. A. Brakhage, U. Horn and C. Hertweck, *ChemBioChem*, 2011, **12**, 1836–1839.
- 251 N. Bohni, M. L. Cordero-Maldonado, J. Maes, D. Siverio-Mota, L. Marcourt, S. Munck, A. R. Kamuhabwa, M. J. Moshi, C. V. Esguerra, P. A. M. de Witte, A. D. Crawford and J. L. Wolfender, *PLoS One*, 2013, **8**, e64006.
- 252 M. Lambert, J. L. Wolfender, D. Stärk, S. B. Christensen, K. Hostettmann and J. W. Jaroszewski, *Anal. Chem.*, 2007, **79**, 727–735.
- 253 J. C. Nielsen, S. Grijseels, S. Prigent, B. Ji, J. Dainat, K. F. Nielsen, J. C. Frisvad, M. Workman and J. Nielsen, *Nat. Microbiol.*, 2017, **2**, 17044.



- 254 N. K. N. Neumann, S. M. Lehner, B. Kluger, C. Bueschl, K. Sedelmaier, M. Lemmens, R. Krska and R. Schuhmacher, *Anal. Chem.*, 2014, **86**, 7320–7327.
- 255 S. Bertrand, O. Schumpp, N. Bohni, M. Monod, K. Gindro and J. L. Wolfender, *J. Nat. Prod.*, 2013, **76**, 1157–1165.
- 256 W. J. Moree, V. V Phelan, C. H. Wu, N. Bandeira, D. S. Cornett, B. M. Duggan and P. C. Dorrestein, *Proc Natl Acad Sci U S A*, 2012, **109**, 13811–13816.