HIGHLIGHT

Check for updates

Biosynthetic gene clusters and the evolution of fungal chemodiversity

Antonis Rokas, [ID] *[a] Matthew E. Mead, [ID] [a] Jacob L. Steenwyk, [ID] [a] Huzefa A. Raja [ID] [b] and Nicholas H. Oberlies [ID] [b]

Covering: up to 2019

Fungi produce a remarkable diversity of secondary metabolites: small, bioactive molecules not required for growth but which are essential to their ecological interactions with other organisms. Genes that participate in the same secondary metabolic pathway typically reside next to each other in fungal genomes and form biosynthetic gene clusters (BGCs). By synthesizing state-of-the-art knowledge on the evolution of BGCs in fungi, we propose that fungal chemodiversity stems from three molecular evolutionary processes involving BGCs: functional divergence, horizontal transfer, and *de novo* assembly. We provide examples of how these processes have contributed to the generation of fungal chemodiversity, discuss their relative importance, and outline major, outstanding questions in the field.

1. Fungal biosynthetic gene clusters produce diverse secondary metabolites of broad ecological importance and human relevance

Fungi produce a remarkable diversity of secondary metabolites,[1] also known as natural products, such as the immunosuppressant cyclosporin,[2] the cholesterol reducing lovastatin,[3] the antibiotic penicillin,[4] the hallucinogenic prodrug psilocybin,[5] and the trichothecene[6] and aflatoxin[7] mycotoxins (Fig. 1). Although these small molecules are not required for

[a]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA. E-mail: antonis.rokas@vanderbilt.edu

[b]Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, USA

fungal survival and growth, their bioactive properties render them highly relevant to human affairs as drugs, toxins, and pigments. But arguably their raison d'être is to act as crucial intermediaries at the front line of fungal ecology. Numerous secondary metabolites are thought to play key roles in shaping the interactions that fungi have with other organisms across the tree of life, including with other fungi,[8] bacteria,[9,10] plants,[11,12] or animals.[13–15] These interactions are varied, and include virulence, defense, quorum sensing, protection, nutrient acquisition and the promotion of growth (Fig. 2).

Most fungal secondary metabolites are encoded by biosynthetic gene clusters (BGCs; Fig. 1); each cluster typically contains the majority, if not all, of the genes participating in the production of a given secondary metabolite, with these genes located adjacent to each other (*i.e.*, "clustered") in the genome.[1,17–19] A typical fungal BGC contains one or more genes

One of the main research foci of the Rokas Research Group at Vanderbilt University (http://www.rokaslab.org) is to understand the molecular basis of fungal metabolic pathway evolution. We do so by studying Aspergillus molds as a model for understanding the molecular evolution of fungal secondary metabolism and Saccharomycotina budding yeasts as a model for understanding the evolution of fungal primary metabolism. Pictured (left to right) are Dr Matthew E. Mead (postdoctoral fellow studying the evolution of complex fungal traits, such as pathogenicity and secondary metabolism), Jacob L. Steenwyk (graduate student working on the dynamics of fungal genome evolution), and Professor Antonis Rokas. Photo credit, Matt Sachs/Genetics Society of America.
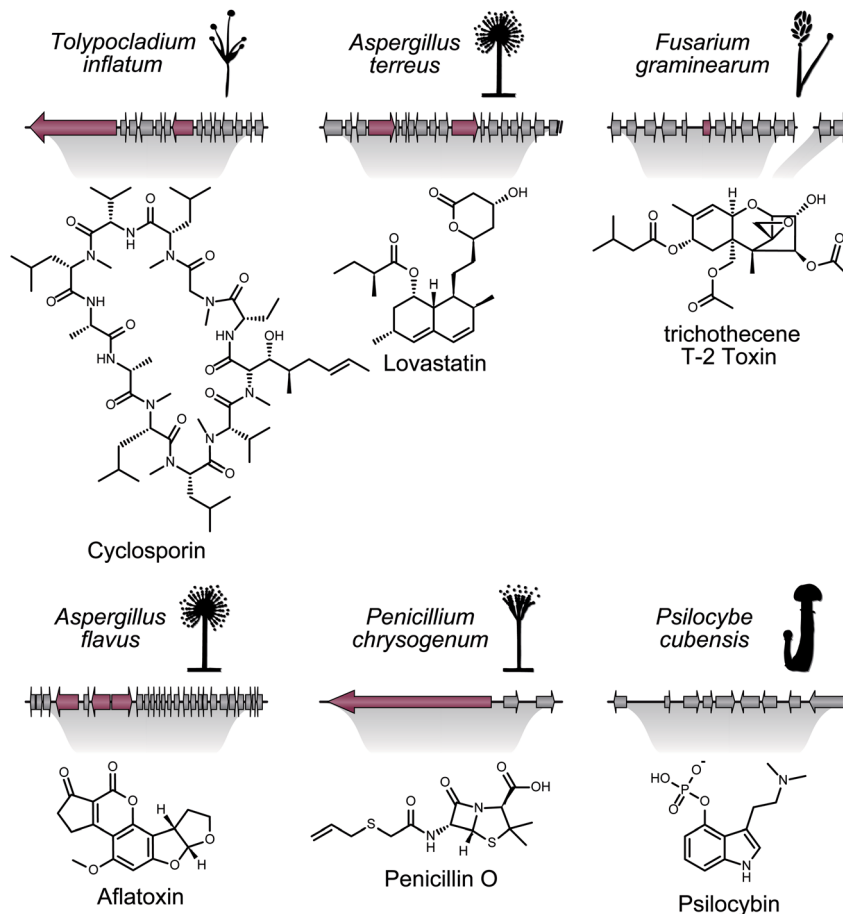
Fig. 1 Select examples of fungal BGCs, their secondary metabolites, and the organisms that produce them. Genes are represented by arrows; genes colored maroon denote secondary metabolite backbone biosynthesis genes (such as polyketide synthases, terpene synthases, and non-ribosomal peptide synthases), whereas genes colored grey denote BGC genes with diverse functions, such as metabolite modification, metabolite transport, regulation of BGC expression, and resistance to secondary metabolite activity. Note that psilocybin biosynthesis does not require any of the canonical backbone biosynthesis genes. Data from: cyclosporin BGC,[2] lovastatin BGC,[3] trichothecene T-2 toxin BGC,[6] aflatoxin BGC,[7] penicillin BGC,[4] and psilocybin BGC.[5,16]

whose protein products catalyze the synthesis of the backbone of the metabolite (such as polyketide synthases, non-ribosomal peptide synthases, and terpene synthases), and one or more genes encoding for: (i) enzymes (such as epimerases, methyl-transferases, and hydroxylases) that modify this backbone, (ii) proteins involved in metabolite transport, (iii) transcription factors involved in regulation of BGC expression, and (iv) proteins that confer resistance to the activity of the secondary metabolite.[1,17] Fungal BGCs are generally similar in their genomic organization to bacterial BGCs; the key difference is

The Oberlies Research Group at the University of North Carolina at Greensboro (https://chem.uncg.edu/oberlies/) works on the isolation and structure elucidation of bioactive compounds from nature, with a particular emphasis on secondary metabolites from fungi. Pictured (left to right) are Dr Huzefa Raja (a Research Scientist and Mycologist) and Professor Nicholas H. Oberlies.
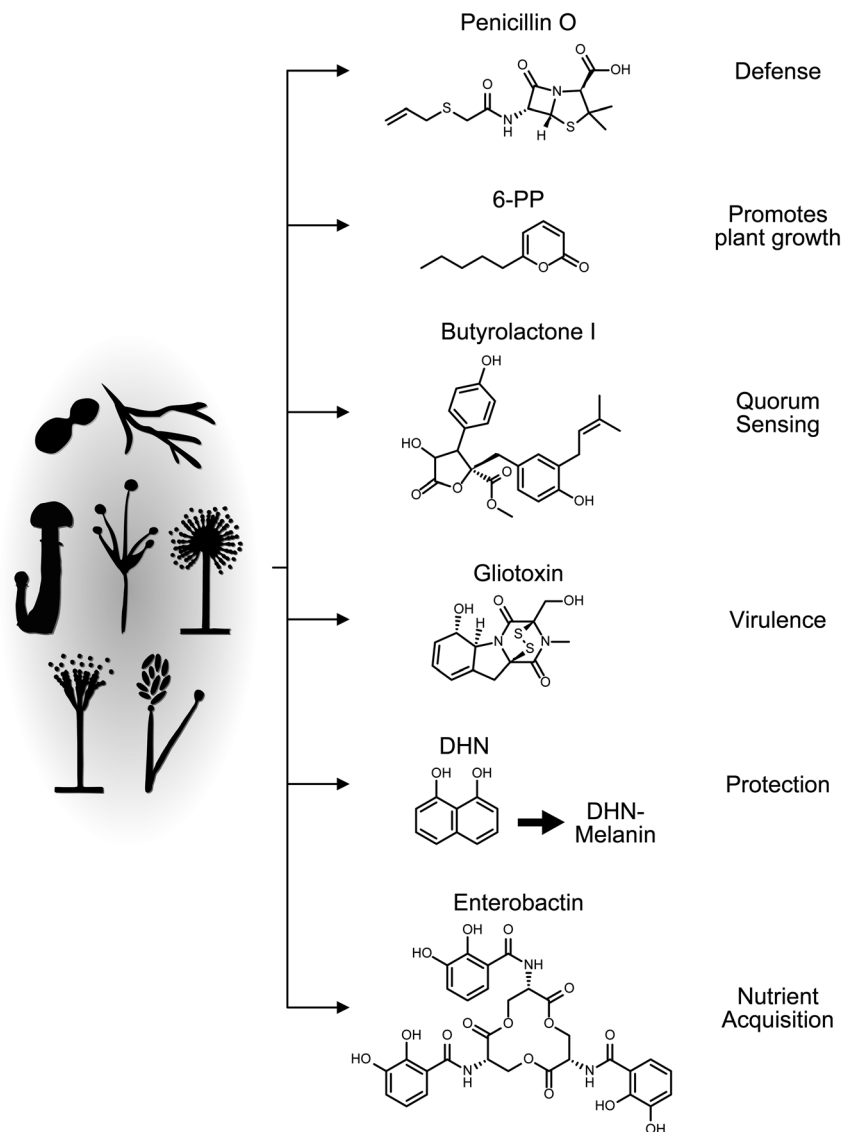
Fig. 2 Secondary metabolites are central to the ecology of many fungi and shape their diverse interactions with other organisms. Penicillin is an antibiotic whose ecological role lies in fungal defense against bacteria,[10] 6-*n*-pentyl-6*H*-pyran-2-one (6-PP) promotes plant growth,[12] butyrolactone I is a quorum sensing molecule,[30] gliotoxin is a virulence factor,[31] DHN-melanin protects again UV light damage,[32] and enterobactin is an iron uptake molecule that contributes to the acquisition of nutrients.[33]

that bacterial BGCs are typically organized into operons (where multiple genes are transcribed into a single messenger RNA), whereas fungal BGCs are typically transcribed individually.[20,21]

Notable secondary metabolites produced by diverse backbone biosynthesis genes and BGCs include: cyclosporin, a non-ribosomal peptide biosynthesized by a 14-gene BGC in the ascomycete fungus *Tolypocladium inflatum*;[2] lovastatin, a polyketide biosynthesized by an 18-gene BGC in the mold *Aspergillus terreus*;[3] the trichothecene T-2 toxin, a terpene biosynthesized by a 12-gene BGC and a 2-gene BGC found in several *Fusarium* species;[6] aflatoxin, a polyketide biosynthesized by a 25-gene BGC in the mold *Aspergillus flavus* and its close relatives;[7] penicillin, a non-ribosomal peptide biosynthesized by a 3-gene BGC in molds in the genera *Penicillium* and *Aspergillus*;[4] and psilocybin, a tryptamine-derived secondary metabolite

biosynthesized by a 9-gene BGC in several different basidiomycete genera whose biosynthesis does not require any of the canonical backbone biosynthesis genes (Fig. 1).[5,16] A comprehensive and up to date compilation of fungal BGCs whose secondary metabolite products have been functionally validated can be found at the MIBiG (Minimum Information about Biosynthetic Gene cluster) repository.[22,23]

BGCs vary widely in their numbers across fungal genomes; whereas ascomycete filamentous fungi and basidiomycete fungi typically contain dozens (if not scores) of BGCs, unicellular yeasts in both lineages either lack BGCs altogether or contain very few.[17,24,25] A given BGC is often known from only a single species or a few closely related ones, but broadly and discontinuously distributed BGCs, such as sterigmatocystin,[26] also exist. Additionally, BGCs and their secondary metabolites also

show extensive variation in their presence/absence patterns within fungal species.[17,27–29]

A notable feature of BGCs, hinted at by their high variability and narrow taxonomic range, is that they are rapidly evolving.[17,34] Why is that so? From a molecular perspective, it has been argued that the lower specificity of secondary metabolic enzymes means that new gene duplicates are more likely to catalyze novel substrates and produce novel products that may be favored by natural selection, accelerating their evolution.[35] Additionally, BGCs often reside in fast-evolving genomic regions, such as near the ends of chromosomes[36] or in accessory chromosomes.[37] From an ecological perspective, the involvement of secondary metabolites in mediating interspecific interactions suggests that they are key in "arms races" between fungi and their competitors, which are thought to accelerate evolutionary rates of the genes involved.[38] But secondary metabolite biosynthesis is also energetically costly. Thus, loss of the ability to produce a secondary metabolite and reliance on other fungal relatives in the community for its production may be, at least up to a point, advantageous to individual organisms[39,40] and further increase the rate of BGC evolution.[17]

One important question raised by considering the ecological relevance of fungal secondary metabolites, the narrowness of their taxonomic distribution, and the fast pace of BGC evolution, concerns the molecular evolutionary processes that give rise to fungal chemodiversity. In this highlight, we suggest that there are three major molecular evolutionary processes that occur at the level of BGCs and which give rise to fungal chemodiversity: functional divergence, horizontal or lateral transfer, and *de novo* assembly (see glossary in Table 1 for definitions of these terms). While the focus of our highlight is on discussing how variation at the level of BGCs gives rise to variation in secondary metabolism or chemodiversity, we note that all genetic variation at the level of BGCs occurs *via* the standard battery of mutational types, such as point mutations, insertions, deletions, rearrangements, duplications, and horizontal gene transfer (see glossary in Table 1 for definitions). All of these types of mutations are well established and known to influence fungal genes, genomes, and BGCs.[20,27,29,34,40]

2. The evolutionary processes underlying fungal chemodiversity

2.1 BGC functional divergence

Functional divergence is the process by which the accumulation of molecular differences between evolutionarily related or homologous (see glossary in Table 1) genes and pathways leads to a change in their function or phenotype. In the context of BGCs, functional divergence refers to the accumulation of molecular differences between the gene sequences of homologous BGCs that then give rise to chemical differences in their secondary metabolite products and generate secondary metabolite structural diversity. Functional divergence has influenced both the evolution of orthologous (see glossary in Table 1) BGCs that have originated through speciation events as well as paralogous (see glossary in Table 1) BGCs that have originated through duplication events.

2.1.1 Functional divergence of orthologous BGCs. Orthologous BGCs can functionally diverge *via* the accumulation of amino acid differences in the enzymes encoded by BGCs. For example, the chemodiversity of fumonisin mycotoxins among *Fusarium* fungi stems from amino acid sequence variation in a protein encoded by a single gene from the fumonisin BGC (Fig. 3A).[41] Some *Fusarium* species, such as *Fusarium verticillioides*, are known to produce primarily fumonisin B, whereas other species, such as *Fusarium oxysporum*, produce primarily fumonisin C. The only difference in the structures of fumonisin B and C is in the length of their backbones; the fumonisin B

Table 1 Glossary of evolutionary terms

Term	Definition
BGC *de novo* assembly	Refers to the process by which an entire BGC is evolutionarily assembled through the recruitment and relocation of native genes, duplicates of native genes, and horizontally acquired genes
BGC duplication	Refers to the generation of an additional (duplicate) copy of an entire BGC in the genome
BGC functional divergence	Refers to the process by which homologous BGCs, through the accumulation of genetic changes, gradually diverge in their functions (*i.e.*, in the secondary metabolites they produce)
BGC horizontal or lateral transfer	Refers to the process by which an entire BGC from the genome of one organism is transferred and stably integrated into the genome of another through non-reproduction related mechanisms
Deletion	Type of mutation, which stems from the deletion of genetic material in the genome
Duplication	Refers to the generation of an additional (duplicate) copy of genetic material in the genome
Homology/homologous	In the context of genes, two genes are homologous if their origins can be traced to the same common ancestor. Homologous genes can originate *via* processes such as gene duplication (in which case they are paralogs) and vertical descent/speciation (in which case they are orthologs)
Horizontal/lateral gene transfer	Refers to the transfer and integration of genetical material from the genome of one organism to the genome of another through non-reproduction related mechanisms
Insertion	Type of mutation, which stems from the insertion of genetic material in the genome
Orthology/orthologous	In the context of genes, two genes are orthologous if they originated *via* vertical descent/speciation, *i.e.*, if they stem from the same ancestral gene that was present in the last common ancestor of the strains/species being compared
Paralogy/paralogous	In the context of genes, two genes are paralogous if they originated *via* gene duplication
Point mutation	Type of mutation, which stems from the replacement of one nucleotide base pair by another
Rearrangement	Type of mutation, which stems from the rearrangement of genetic material in the genome

A

Fumonisin B

Fusarium verticillioides

Fusarium oxysporum

Complementation of *FUM8*

Fumonisin C

Fumonisin B

Fumonisin C

| ▭ | F. verticillioides FUM8 ortholog | ▭ | F. oxysporum FUM8 ortholog | ▭ | Other FUM genes |

B

Aflatoxin

Aspergillus flavus

aflF aflU aflT   aflC   aflD   aflB   aflA   aflR aflS aflH aflJ aflE aflM aflN aflG aflL aflI aflO aflP aflQ aflK aflV aflW aflX aflY

Aspergillus nidulans

stcA stcB   stcC stcD stcE aflR aflJ stcF stcG stcI   stcJ   stcK   stcL stcM stcN stcO stcQ stcR stcS stcT stcU stcV stcW

Sterigmatocystin

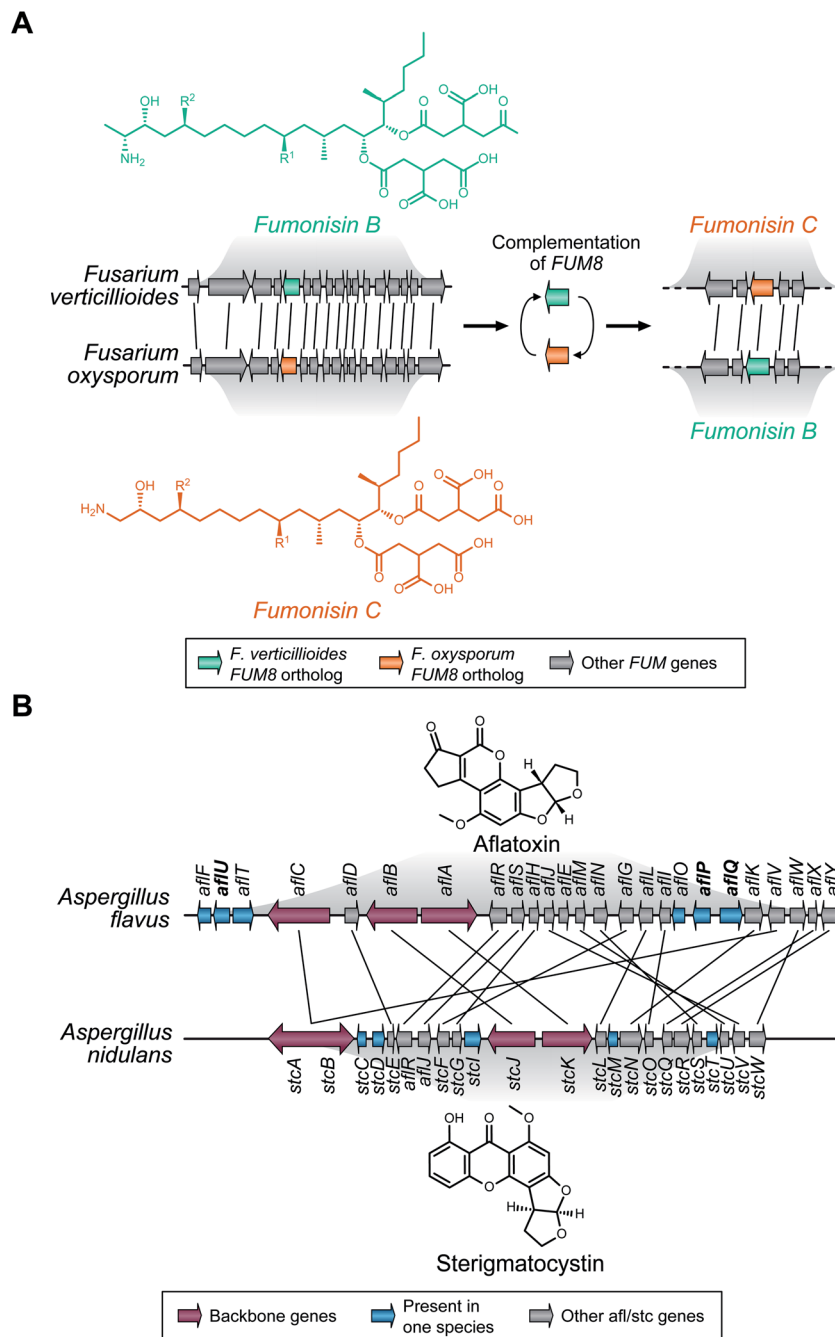| ▭ | Backbone genes | ▭ | Present in one species | ▭ | Other afl/stc genes |

Fig. 3 Fumonisins (panel A) and sterigmatocystin/aflatoxins (panel B); two notable examples of fungal chemodiversity that stems from the functional divergence of orthologous BGCs. Genes are represented by arrows. Lines between genes from different species refer to orthologous genes.

backbone is 20 carbon atoms long, whereas the backbone of fumonisin C is 19 carbon atoms long. Comparison of the fumonisin BGCs in *F. verticillioides* and *F. oxysporum* showed that the two species contain orthologous BGCs with the same 19 (orthologous) genes; gene swapping experiments further showed that sequence variation within the *fum8* gene, which encodes for an α-oxoamine synthase, is responsible for the observed difference in the type of fumonisin (B or C) produced by the two species.[41] The precise amino acid difference(s)

between the *F. verticillioides* and *F. oxysporum* Fum8 protein orthologs responsible for observed divergence in fumonisin structure are not known and the two orthologs exhibit 91% similarity in their amino acid sequences.[41] However, it appears that the *F. verticillioides* Fum8 enzyme preferentially binds the amino acid alanine (and catalyzes its condensation to an 18-carbon linear polyketide to produce the 20-carbon-long fumonisin B), whereas the *F. oxysporum* Fum8 preferentially binds glycine, resulting in the production of the 19-carbon-long

fumonisin C.[41] Sequence comparisons of *Fusarium* α-oxoamine synthase sequences show that the amino acid residue at position 580 of the protein is strongly associated with the type of fumonisin produced; presence of alanine at position 580 is associated with fumonisin B production, whereas presence of valine at the same position is associated with fumonisin C production.[42] Consistent with this association, mutations of this residue in human α-oxoamine synthase have been shown to alter the enzyme's binding affinity to its amino acid substrate.[43]

Alternatively, orthologous BGCs can functionally diverge through gains and losses of genes (Fig. 3B). For example, some *Aspergillus* species, such as *Aspergillus flavus*, produce the mycotoxin aflatoxin, whereas other species, including *Aspergillus nidulans*, produce the mycotoxin sterigmatocystin. The two mycotoxins, as well as their BGCs, are similar to each other. The difference in the mycotoxin produced is due to at least three genes (*aflP*, *aflU*, and *aflQ*; shown in bold in Fig. 3B) present in the aflatoxin BGC that are not found in the sterigmatocystin BGC. The AflP protein is an *O*-methyltransferase that converts sterigmatocystin to *O*-methylsterigmatocystin, whereas the cytochrome P450 monooxygenase AflU and the P-450 monooxygenase AflQ catalyze the conversion of *O*-methylsterigmatocystin to aflatoxin G and aflatoxin B, respectively.[7,44] Note that the differences in gene content between the aflatoxin and sterigmatocystin BGCs include additional genes (Fig. 3B); however, only *aflP*, *aflU*, and *aflQ* have been shown to be involved in the conversion of sterigmatocystin to the aflatoxins.

Finally, some orthologous BGCs have functionally diverged through both the accumulation of amino acid differences in the protein products of their genes as well as through gains and losses of genes. The combined effect of these two processes is thought to account for the observed structural diversity of yanuthone antimicrobial compounds in *Penicillium* molds,[45] as well as for the diversity of the echinocandin class of antifungal drugs,[46] trichothecene mycotoxins,[47] and ergot alkaloids[48] produced by diverse fungi.

2.1.2 Functional divergence of paralogous BGCs. BGC functional divergence that gives rise to the evolution of new secondary metabolites can also occur *via* the duplication of genomic regions containing entire BGCs. Even though duplication of genes in BGCs has been widely documented and it is now well established that gene duplication is a major driver of both the diversity of individual backbone genes present in BGCs[49,50] as well as of genes in BGCs in general,[34] much less is known about the duplication of entire BGCs.

An example of BGC duplication concerns the duplication of two polyketide-producing BGCs, *Pks1-gc* and *Pks2-gc*, in *Metarhizium* entomopathogenic fungi, one of which is known to produce an anthraquinone derivative.[51] Genomic and functional analyses of the two paralogous BGCs show that they have functionally diverged through the reciprocal loss of genes in each BGC as well as through the accumulation of substitutions in both the promoter and protein-coding regions of their polyketide synthase genes (Fig. 4).[51] Interestingly, the only shared paralogous gene pair between the *Pks1-gc* and the *Pks2-gc* is the *Pks1–Pks2* pair. In contrast, the *Pks1-gc* and the *Pks2-gc* BGCs share two and three homologous genes, respectively, with the *A. fumigatus* conidial pigment BGC (Fig. 4). Consistent with these differences in gene sequence and content, the two BGCs show distinct expression patterns (the genes of *Pks1-gc* are expressed during asexual spore formation, whereas the genes of *Pks2-gc* are expressed during the establishment of infection in insects) and produce distinct secondary metabolites. The anthraquinone derivative product of *Pks1-gc* is involved in the pigmentation of asexual spores and in abiotic stress tolerance, such as tolerance to UV light, whereas the uncharacterized product of *Pks2-gc* appears to contribute to pathogenicity and not pigmentation or abiotic stress.[51]
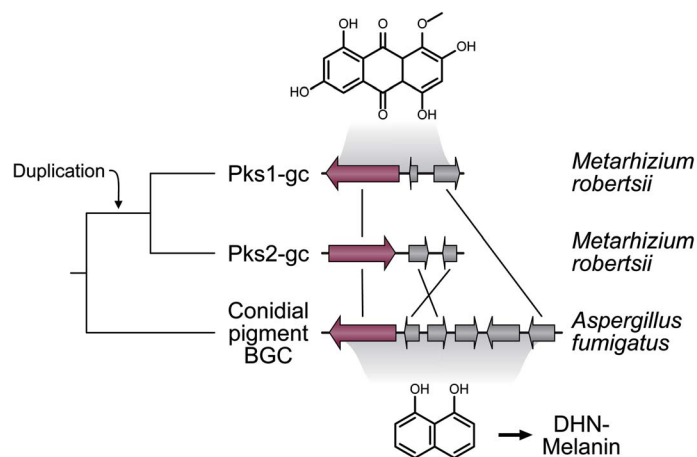

Fig. 4 Duplication and subsequent functional divergence of *Pks1-gc* and *Pks2-gc*, two paralogous polyketide BGCs present in *Metarhizium* insect pathogens. Two polyketide synthase-containing BGCs, *Pks1-gc* and *Pks2-gc*, in *Metarhizium robertsii* appear to be the result of an ancient duplication of an entire BGC that likely resembled the conidial pigment BGC found in the mold *Aspergillus fumigatus*. The *Pks1-gc* BGC produces an anthraquinone derivative, whereas the product of the *Pks2-gc* has yet to be characterized.[51] Data from: *M. robertsii Pks1-gc* BGC[51] and chemical structure; *M. robertsii Pks2-gc* BGC;[51] and *A. fumigatus* conidial pigment BGC.[52] Lines between genes from different species refer to orthologous genes.

This journal is © The Royal Society of Chemistry 2019

Another example of functional divergence of paralogous BGCs are the patulin and yanuthone BGCs in *Penicillium* molds, whose secondary metabolite products share a 6-methylsalicylic acid (6-MSA) core.[45] The 15-gene patulin BGC and the 10-gene yanuthone BGCs contain several pairs of paralogous genes thought to catalyze the same reactions leading to the formation of the 6-MSA core structure as well as several additional genes that lack sequence similarity to genes in the other BGC and are presumably responsible for the structural differences between patulin and yanuthones.[45] Thus, a proto-BGC responsible for the production of 6-MSA likely originated and duplicated prior to the origin of *Penicillium*, followed by additional recruitment of non-homologous genes in both BGCs. Interestingly, phylogenetic analysis of the 6-MSA synthase protein suggests that the patulin 6-MSA synthase is more closely related to the 6-MSA synthases found in the aculinic acid BGC from *Aspergillus aculeatus*[53] and in the terreic acid BGC in *Aspergillus terreus*,[54] both of which produce 6-MSA-based secondary metabolites.[45] Thus, the duplication and subsequent functional divergence of the patulin and yanuthone BGCs is part of a broader series of duplication and functional divergence events of 6-MSA-based BGCs.

2.2 BGC horizontal transfer

Fungal chemodiversity can also originate *via* the horizontal transfer of entire BGCs from other organisms.[55] For example, horizontal transfer of the sterigmatocystin BGC from *Aspergillus* to *Podospora* resulted in the ability of the latter to produce sterigmatocystin (Fig. 5).[56,57] In the aftermath of horizontal transfer, the acquired BGCs can accumulate changes in their sequence and genomic organization without altering the structure of the metabolic product. For example, the average amino acid sequence similarity between the proteins encoded by the *Aspergillus nidulans* and *Podospora anserina* sterigmatocystin BGCs is 63% and the two BGCs also differ somewhat in their genomic organization, yet both produce the same metabolite. Thus, in contrast to BGC functional divergence (Section 2.2) and BGC *de novo* assembly (Section 2.4 below), both of which result in BGCs that produce new compounds, BGC horizontal transfer typically results in the production of an existing compound in a new, typically distantly related, organism.

In the last decade, several examples of BGC horizontal transfer have been reported; most transfers of entire BGCs are between fungi, such as the transfers of the BGC for the pigment bikaverin from the ascomycete genus *Fusarium* to that of *Botrytis*,[58–60] of the BGC for the hallucinogen psilocybin among basidiomycete fungi,[16] of the fumonisin BGC across *Fusarium* species,[42] of the chaetoglobosin-like BGC from *Penicillium* to *Mycosphaerella populorum*,[61] or the multiple transfers of the BGC for the histone deacetylase inhibitor depudecin among ascomycete fungi.[62] In contrast, horizontal transfer of entire BGCs from bacteria, the lineage in which secondary metabolism first originated,[63] appears to be less common and only one clear-cut example of transfer of the siderophore enterobactin from enterobacteria to budding yeasts is known to date.[33]
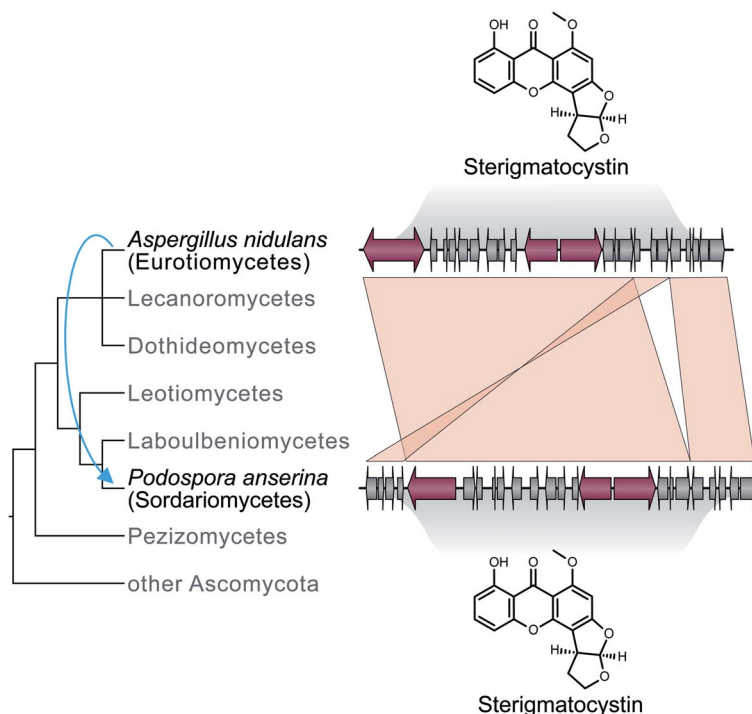

Fig. 5 Horizontal transfer of the sterigmatocystin BGC from *Aspergillus* to *Podospora* resulted in the presence of the sterigmatocystin BGC in the *Podospora* genome and its ability to produce sterigmatocystin. Evolutionary analyses of the history of the genes in the sterigmatocystin BGC suggest that the *Podospora* BGC was horizontally acquired from an *Aspergillus* ancestor.[56] Subsequent functional and chemical studies have validated that *Podospora* fungi produce the sterigmatocystin mycotoxin.[10,57,65] Large orthologous blocks of genetic sequence are depicted using orange trapezoids.

The examples discussed above all concern transfers of BGCs in the absence of functional divergence (*i.e.*, the same secondary metabolite is produced in both the donor and the recipient organisms). The identification of examples of BGCs that functionally diverged after HGT is more challenging because, following functional divergence, donor and recipient BGCs can exhibit substantial divergence in gene content and arrangement.[55] BGC horizontal transfer followed by functional divergence is thought to account for the diversification of epipolythiodioxopiperazine (ETP) mycotoxins, such as gliotoxin, sirodesmin and their relatives.[64]

2.3   *De novo* BGC assembly

The final, and least well-documented, evolutionary process involved in the generation of fungal chemodiversity is *de novo* BGC assembly, under which new secondary metabolites originate from scratch in fungal genomes. The genes that become part of the newly formed secondary metabolic pathway originate either through duplication and relocation of native genes or through horizontal acquisition. One important source of genes for BGCs are duplicates of genes encoding for enzymes already involved in primary and secondary metabolism, such as the isopropyl-malate synthase gene in the echinocandin BGC of *Aspergillus rugulosus*,[66] and the citrate synthase gene in the zaragozic acid BGC of *Curvularia lunata*.[67]

De novo assembled BGCs are unlikely to be highly similar in their gene or sequence content to already existing BGCs, making their identification through comparisons of genome sequences (the major way all cases of BGC functional divergence and BGC horizontal transfer have been identified) much more challenging. Several lines of evidence support that this mechanism also gives rise to fungal BGCs. The same general evolutionary process of *de novo* pathway assembly is thought to be responsible for the origin of novel pathways that break down anthropogenic chemicals[68] as well as of certain catabolic pathways.[11,69]

De novo secondary metabolic pathways may originate in a similar manner *via* a two-step process; step one involves the assembly of the secondary metabolic pathway through the recruitment of native genes, duplicates of native genes, and horizontally acquired genes, and step two involves their clustering into a BGC. Consistent with this model, several fungal secondary metabolic pathways are comprised of two or more BGCs,[17] suggesting that the clustering of fungal secondary metabolic pathways is not an absolute requirement for their function. For example, a 12-gene and a 2-gene BGC found in distinct genomic locations are involved in the biosynthesis of the trichothecene T-2 toxin in *F. graminearum* (Fig. 1).[6] Additionally, several BGCs contain distinct smaller clusters of genes (often referred to as modules) responsible for the production of functional intermediates within the pathway, suggesting that the entire BGC evolved *via* the merging of distinct, pre-existing smaller BGCs. For example, BGCs associated with the production of echinocandins typically contain a 4-gene cluster for the production of L-homotyrosine, one of the intermediates required for echinocandin biosynthesis.[46] Similarly, the genes of BGCs responsible for the production of distinct secondary metabolites

can be intertwined in the genome, as in the case of the fumagillin and pseurotin BGCs in *A. fumigatus*,[70] providing empirical evidence of the evolutionary merging of distinct BGCs.

The second line of evidence is that gene relocation has been implicated in the diversification of BGCs, such as the expansion of a trichothecene BGC in *Fusarium* species.[71] Perhaps the best candidate of *de novo* assembly of a BGC involved in secondary metabolism is the fumonisin BGC found in certain *Fusarium* and *Aspergillus* species.[72] While the presence of the BGC in *Aspergillus* is best explained by horizontal transfer from another fungus, one hypothesis for the origin of the *Fusarium* BGC, based on examination of phylogenies of genes in the BGC, is that it arose through the relocation and clustering of genes that were originally dispersed in the genome.[72]

3.   Perspective and major unanswered questions

Even though the remarkable breadth of fungal chemodiversity was well appreciated before the advent of the genomics revolution,[73] the sequencing of diverse fungal genomes from 2003 onward quickly began revealing that fungal genomes contained even larger numbers of BGCs responsible for the biosynthesis of yet-unknown secondary metabolite products and provided unprecedented opportunities for studying the origins and evolution of fungal chemodiversity at the DNA sequence level.[17] Currently, the molecular evolutionary processes by which fungal BGCs evolve are becoming established (Fig. 3) and the relationship between chemical diversity and BGC diversity for several secondary metabolites is being increasingly refined.[47]

Comparison of the genetic and evolutionary mechanisms underpinning the evolution of fungal chemodiversity with those inferred from the study of bacterial chemodiversity[74,75] suggests that similar mechanisms operate in both lineages. Arguably the biggest difference is the extent of the contribution of BGC horizontal transfer in driving chemodiversity in the two lineages. Although the role of BGC horizontal transfer is increasingly appreciated in fungi (see Section 2.3), bacterial BGC horizontal transfer occurs at far higher rates and plays a bigger role in shaping bacterial chemodiversity.[74]

While the major contours of the molecular evolutionary basis of fungal chemodiversity are increasingly well understood, several major outstanding questions and opportunities remain. For example, we still lack an understanding of why fungal secondary metabolic pathways are typically arranged in the genome as BGCs (three genetic models, namely co-regulation, genetic linkage, and selfishness, and one phenotypic model, namely toxicity avoidance, have been put forward as explanations)[17,19,76] and whether this clustering is associated with fungal chemodiversity. We similarly lack a complete knowledge of the distribution and genomic arrangement of secondary metabolic pathways in fungal genomes, especially from less-studied and less-sequenced lineages located outside a few select genera of filamentous fungi (*e.g.*, *Aspergillus*, *Fusarium*, *Penicillium*) from the phylum Ascomycota.[77]

In the context of this highlight article, arguably the biggest challenges and opportunities lie in uncovering examples of *de novo* BGC assembly, understanding the relative contribution of the three different processes in sculpting BGC diversity, and elucidating how this diversity translates to chemodiversity. Recently developed computational algorithms now allow the construction of networks of fungal BGCs on the basis of their sequence similarity and gene order, enabling the grouping of BGCs into BGC families, of families into clans, and so on.[78] Reconciling this network view of BGC evolution with the evolutionary processes that we discuss promises to illuminate their relative importance in sculpting BGC diversity and how that translated to chemodiversity. For example, a recent examination of 37 *Aspergillus* and *Penicillium* genomes identified more than 2700 BGCs that could be grouped into 455 BGC families that presumably produce distinct groups of secondary metabolites; strikingly, nearly half of these families contained only a single BGC.[79] How did these single-BGC families originate and how common are they when the entirety of fungal genomes is examined? And how do these 455 BGC families relate to the ~15 600 described fungal secondary metabolites?[10] These are exciting questions but also non-trivial to address, not least because of the challenges associated with handling and analyzing the ever increasing volume of publicly available fungal genomes (there are 5064 draft fungal genomes in Gen-Bank as of October 30, 2019).

But the opportunity does not stop here; by considering the mechanisms that give rise to BGC diversity we begin to set the foundations of an evolutionary framework to bridge genotype (BGCs) with chemotype (their secondary metabolites). Establishing such a framework will not only advance our understanding of how genomic diversity translates to chemodiversity, but will also be useful in genetic engineering- and directed evolution-based efforts to discover and produce new leads in the pharmaceutical and agrochemical research areas.[45] Connecting BGC diversity with chemodiversity, and elucidating the relationship between BGC sequence divergence and chemical structure divergence, is even more daunting due to the current lack of structures for most fungal BGCs[17] and *vice versa* (*i.e.*, the BGCs responsible for making most fungal secondary metabolites are unknown). With the sequences of tens of thousands of fungal BGCs, thousands of fungal secondary metabolite chemical structures, and a smorgasbord of novel synthetic biology, chemical, and bioinformatic tools that accelerate the discovery of new secondary metabolites[80] at hand, exciting discoveries lay ahead.

4. Conflicts of interest

There are no conflicts of interest to declare.

5. Acknowledgements

6. References

1 N. P. Keller, *Nat. Rev. Microbiol.*, 2019, **17**, 167–180.
2 K. E. Bushley, R. Raja, P. Jaiswal, J. S. Cumbie, M. Nonogaki, A. E. Boyd, C. A. Owensby, B. J. Knaus, J. Elser, D. Miller, Y. Di, K. L. McPhail and J. W. Spatafora, *PLoS Genet.*, 2013, **9**, e1003496.
3 K. C. Mulder, F. Mulinari, O. L. Franco, M. S. Soares, B. S. Magalhaes and N. S. Parachin, *Biotechnol. Adv.*, 2015, **33**, 648–665.
4 S. Gutierrez, F. Fierro, J. Casqueiro and J. F. Martin, *Antonie van Leeuwenhoek*, 1999, **75**, 81–94.
5 J. Fricke, F. Blei and D. Hoffmeister, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 12352–12355.
6 N. J. Alexander, R. H. Proctor and S. P. McCormick, *Toxin Rev.*, 2009, **28**, 198–215.
7 J. Yu, P. K. Chang, K. C. Ehrlich, J. W. Cary, D. Bhatnagar, T. E. Cleveland, G. A. Payne, J. E. Linz, C. P. Woloshuk and J. W. Bennett, *Appl. Environ. Microbiol.*, 2004, **70**, 1253–1262.
8 S. L. Knowles, H. A. Raja, A. J. Wright, A. M. L. Lee, L. K. Caesar, N. B. Cech, M. E. Mead, J. L. Steenwyk, L. N. A. Ries, G. H. Goldman, A. Rokas and N. H. Oberlies, *Front. Microbiol.*, 2019, **10**, 285.
9 E. B. Van Arnam, C. R. Currie and J. Clardy, *Chem. Soc. Rev.*, 2018, **47**, 1638–1651.
10 G. F. Bills and J. B. Gloer, *Microbiol. Spectrum*, 2016, **4**, FUNK-0009-2016.
11 E. Gluck-Thaler, V. Vijayakumar and J. C. Slot, *Mol. Ecol.*, 2018, **27**, 5120–5136.
12 F. Vinale, K. Sivasithamparam, E. L. Ghisalberti, R. Marra, M. J. Barbetti, H. Li, S. L. Woo and M. Lorito, *Physiol. Mol. Plant Pathol.*, 2008, **72**, 80–86.
13 M. T. Drott, T. Debenport, S. A. Higgins, D. H. Buckley and M. G. Milgroom, *mBio*, 2019, **10**, e02782-18.
14 S. Caballero Ortiz, M. Trienens, K. Pfohl, P. Karlovsky, G. Holighaus and M. Rohlfs, *Ecol. Evol.*, 2018, **8**, 4328–4339.
15 M. Rohlfs, *Front. Microbiol.*, 2014, **5**, 788.
16 H. T. Reynolds, V. Vijayakumar, E. Gluck-Thaler, H. B. Korotkin, P. B. Matheny and J. C. Slot, *Evolution Letters*, 2018, **2**, 88–101.
17 A. Rokas, J. H. Wisecaver and A. L. Lind, *Nat. Rev. Microbiol.*, 2018, **16**, 731–744.
18 J. C. Slot, *Adv. Genet.*, 2017, **100**, 141–178.
19 H. W. Nutzmann, C. Scazzocchio and A. Osbourn, *Annu. Rev. Genet.*, 2018, **52**, 159–183.
20 A. Osbourn, *Trends Genet.*, 2010, **26**, 449–457.

21 Q. Yue, L. Chen, Y. Li, G. F. Bills, X. Zhang, M. Xiang, S. Li, Y. Che, C. Wang, X. Niu, Z. An and X. Liu, *mBio*, 2015, **6**, e00703-15.

22 Y. F. Li, K. J. S. Tsai, C. J. B. Harvey, J. J. Li, B. E. Ary, E. E. Berlew, B. L. Boehman, D. M. Findley, A. G. Friant, C. A. Gardner, M. P. Gould, J. H. Ha, B. K. Lilley, E. L. McKinstry, S. Nawal, R. C. Parry, K. W. Rothchild, S. D. Silbert, M. D. Tentilucci, A. M. Thurston, R. B. Wai, Y. Yoon, R. S. Aiyar, M. H. Medema, M. E. Hillenmeyer and L. K. Charkoudian, *Fungal Genet. Biol.*, 2016, **89**, 18–28.

23 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. Dusterhus, 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. Kotter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H. W. Nutzmann, G. H. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. H. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. H. Yan, G. Yim, F. G. Yu, Y. C. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gomez, 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. Hofte, S. E. Jensen, J. H. 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. Linquan, X. Y. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Mendez, M. Metsa-Ketela, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Muller, 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. Sussmuth, 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. S. Zhang, S. B. Zotchev, R. Breitling, E. Takano and F. O. Glockner, *Nat. Chem. Biol.*, 2015, **11**, 625–631.

24 D. J. Krause, J. Kominek, D. A. Opulente, X. X. Shen, X. Zhou, Q. K. Langdon, J. DeVirgilio, A. B. Hulfachor, C. P. Kurtzman, A. Rokas and C. T. Hittinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 11030–11035.

25 S. E. Helaly, B. Thongbai and M. Stadler, *Nat. Prod. Rep.*, 2018, **35**, 992–1014.

26 C. Rank, K. F. Nielsen, T. O. Larsen, J. Varga, R. A. Samson and J. C. Frisvad, *Fungal Biol.*, 2011, **115**, 406–420.

27 A. L. Lind, J. H. Wisecaver, C. Lameiras, P. Wiemann, J. M. Palmer, N. P. Keller, F. Rodrigues, G. H. Goldman and A. Rokas, *PLoS Biol.*, 2017, **15**, e2003583.

28 F. E. Hartmann and D. Croll, *Mol. Biol. Evol.*, 2017, **34**, 2808–2822.

29 R. A. Olarte, J. Menke, Y. Zhang, S. Sullivan, J. C. Slot, Y. Huang, J. P. Badalamenti, A. C. Quandt, J. W. Spatafora and K. E. Bushley, *MC Genomics*, 2019, **20**, 120.

30 T. G. Schimmel, A. D. Coffman and S. J. Parsons, *Appl. Environ. Microbiol.*, 1998, **64**, 3707–3712.

31 J. A. Sugui, J. Pardo, Y. C. Chang, K. A. Zarember, G. Nardone, E. M. Galvez, A. Mullbacher, J. I. Gallin, M. M. Simon and K. J. Kwon-Chung, *Eukaryotic Cell*, 2007, **6**, 1562–1569.

32 C. Kawamura, T. Tsujimoto and T. Tsuge, *Mol. Plant-Microbe Interact.*, 1999, **12**, 59–63.

33 J. Kominek, D. T. Doering, D. A. Opulente, X. X. Shen, X. Zhou, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, M. A. McGee, S. D. Karlen, C. P. Kurtzman, A. Rokas and C. T. Hittinger, *Cell*, 2019, **176**, 1356–1366.

34 J. H. Wisecaver, J. C. Slot and A. Rokas, *PLoS Genet.*, 2014, **10**, e1004816.

35 J. K. Weng, R. N. Philippe and J. P. Noel, *Science*, 2012, **336**, 1667–1670.

36 J. M. Palmer and N. P. Keller, *Curr. Opin. Microbiol.*, 2010, **13**, 431–436.

37 K. Hoogendoorn, L. Barra, C. Waalwijk, J. S. Dickschat, T. A. J. van der Lee and M. H. Medema, *Front. Microbiol.*, 2018, **9**, 1158.

38 S. Paterson, T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall and M. A. Brockhurst, *Nature*, 2010, **464**, 275–278.

39 J. J. Morris, R. E. Lenski and E. R. Zinser, *mBio*, 2012, **3**, e00036-12.

40 T. A. Richards and N. J. Talbot, *Nat. Rev. Microbiol.*, 2013, **11**, 720–727.

41 R. H. Proctor, M. Busman, J. A. Seo, Y. W. Lee and R. D. Plattner, *Fungal Genet. Biol.*, 2008, **45**, 1016–1026.

42 R. H. Proctor, F. Van Hove, A. Susca, G. Stea, M. Busman, T. van der Lee, C. Waalwijk, A. Moretti and T. J. Ward, *Mol. Microbiol.*, 2013, **90**, 290–306.

43 P. M. Shoolingin-Jordan, S. Al-Daihan, D. Alexeev, R. L. Baxter, S. S. Bottomley, I. D. Kahari, I. Roy, M. Sarwar, L. Sawyer and S. F. Wang, *Biochim. Biophys. Acta*, 2003, **1647**, 361–366.

44 J. Yu, *Toxins*, 2012, **4**, 1024–1057.

45 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.

46 Q. Yue, L. Chen, X. Zhang, K. Li, J. Sun, X. Liu, Z. An and G. F. Bills, *Eukaryotic Cell*, 2015, **14**, 698–718.

47 R. H. Proctor, S. P. McCormick, H. S. Kim, R. E. Cardoza, A. M. Stanley, L. Lindo, A. Kelly, D. W. Brown, T. Lee, M. M. Vaughan, N. J. Alexander, M. Busman and S. Gutierrez, *PLoS Pathog.*, 2018, **14**, e1006946.

48 C. A. Young, C. L. Schardl, D. G. Panaccione, S. Florea, J. E. Takach, N. D. Charlton, N. Moore, J. S. Webb and J. Jaromczyk, *Toxins*, 2015, **7**, 1273–1302.

49 B. J. Condon, Y. Leng, D. Wu, K. E. Bushley, R. A. Ohm, R. Otillar, J. Martin, W. Schackwitz, J. Grimwood, N. 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.

50 G. Koczyk, A. Dawidziuk and D. Popiel, *Genome Biol. Evol.*, 2015, **7**, 3132–3154.

51 G. Zeng, P. Zhang, Q. Zhang, H. Zhao, Z. Li, X. Zhang, C. Wang, W. B. Yin and W. Fang, *PLoS Genet.*, 2018, **14**, e1007472.

52 H. F. Tsai, M. H. Wheeler, Y. C. Chang and K. J. Kwon-Chung, *J. Bacteriol.*, 1999, **181**, 6469–6477.

53 L. M. Petersen, D. K. Holm, C. H. Gotfredsen, U. H. Mortensen and T. O. Larsen, *Chembiochem*, 2015, **16**, 2200–2204.

54 C. J. Guo, W. W. Sun, K. S. Bruno and C. C. Wang, *Org. Lett.*, 2014, **16**, 5250–5253.

55 J. H. Wisecaver and A. Rokas, *Front. Microbiol.*, 2015, **6**, 161.

56 J. C. Slot and A. Rokas, *Curr. Biol.*, 2011, **21**, 134–139.

57 L. Shen, F. H. Poree, T. Gaslonde, H. Lalucque, F. Chapeland-Leclerc and G. Ruprich-Robert, *Environ. Microbiol.*, 2019, **21**, 3011–3026.

58 M. A. Campbell, M. Staats, J. L. A. van Kan, A. Rokas and J. C. Slot, *Mycologia*, 2013, **105**, 1126–1134.

59 M. A. Campbell, A. Rokas and J. C. Slot, *Genome Biol. Evol.*, 2012, **4**, 289–293.

60 J. E. Spraker, P. Wiemann, J. A. Baccile, N. Venkatesh, J. Schumacher, F. C. Schroeder, L. M. Sanchez and N. P. Keller, *mBio*, 2018, **9**, e00820-18.

61 B. Dhillon, N. Feau, A. L. Aerts, S. Beauseigle, L. Bernier, A. Copeland, A. Foster, N. Gill, B. Henrissat, P. Herath, K. M. LaButti, A. Levasseur, E. A. Lindquist, E. Majoor, R. A. Ohm, J. L. Pangilinan, A. Pribowo, J. N. Saddler, M. L. Sakalidis, R. P. de Vries, I. V. Grigoriev, S. B. Goodwin, P. Tanguay and R. C. Hamelin, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 3451–3456.

62 H. T. Reynolds, J. C. Slot, H. H. Divon, E. Lysoe, R. H. Proctor and D. W. Brown, *Mol. Biol. Evol.*, 2017, **34**, 2002–2015.

63 T. Cavalier-Smith, *Ciba Found. Symp.*, 1992, **171**, 64–80; discussion 80–67.

64 N. J. Patron, R. F. Waller, A. J. Cozijnsen, D. C. Straney, D. M. Gardiner, W. C. Nierman and B. J. Howlett, *BMC Evol. Biol.*, 2007, **7**, 174.

65 J. C. Matasyoh, B. Dittrich, A. Schueffler and H. Laatsch, *Parasitol. Res.*, 2011, **108**, 561–566.

66 R. A. Cacho, W. Jiang, Y. H. Chooi, C. T. Walsh and Y. Tang, *J. Am. Chem. Soc.*, 2012, **134**, 16781–16790.

67 N. Liu, Y. S. Hung, S. S. Gao, L. Hang, Y. Zou, Y. H. Chooi and Y. Tang, *Org. Lett.*, 2017, **19**, 3560–3563.

68 S. D. Copley, *Nat. Chem. Biol.*, 2009, **5**, 559–566.

69 S. Wong and K. H. Wolfe, *Nat. Genet.*, 2005, **37**, 777–782.

70 P. Wiemann, C. J. Guo, J. M. Palmer, R. Sekonyela, C. C. Wang and N. P. Keller, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 17065–17070.

71 R. H. Proctor, S. P. McCormick, N. J. Alexander and A. E. Desjardins, *Mol. Microbiol.*, 2009, **74**, 1128–1142.

72 N. Khaldi and K. H. Wolfe, *Int. J. Evol. Biol.*, 2011, **2011**, 423821.

73 C. Pearce, *Adv. Appl. Microbiol.*, 1997, **44**, 1–80.

74 A. C. Ruzzini and J. Clardy, *Curr. Biol.*, 2016, **26**, R859–R864.

75 M. A. Fischbach, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4601–4608.

76 K. L. McGary, J. C. Slot and A. Rokas, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 11481–11486.

77 M. Stadler and D. Hoffmeister, *Front. Microbiol.*, 2015, **6**, 127.

78 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. Dias Cappelini, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *bioRxiv*, 2018.

79 T. C. Vesth, J. L. Nybo, S. Theobald, J. C. Frisvad, T. O. Larsen, K. F. Nielsen, J. B. Hoof, J. Brandl, A. Salamov, R. Riley, J. M. Gladden, P. Phatale, M. T. Nielsen, E. K. Lyhne, M. E. Kogle, K. Strasser, E. McDonnell, K. Barry, A. Clum, C. Chen, K. LaButti, S. Haridas, M. Nolan, L. Sandor, A. Kuo, A. Lipzen, M. Hainaut, E. Drula, A. Tsang, J. K. Magnuson, B. Henrissat, A. Wiebenga, B. A. Simmons, M. R. Makela, R. P. de Vries, I. V. Grigoriev, U. H. Mortensen, S. E. Baker and M. R. Andersen, *Nat. Genet.*, 2018, **50**, 1688–1695.

80 C. Greco, N. P. Keller and A. Rokas, *Curr. Opin. Microbiol.*, 2019, **51**, 22–29.