

## Accepted Article

**Title:** Core Steps to the Azaphilone Family of Fungal Natural Products

**Authors:** Katherine Williams, Claudio Greco, Andrew M. Bailey, and  
Christine L. Willis

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

**To be cited as:** *ChemBioChem* 10.1002/cbic.202100240

**Link to VoR:** <https://doi.org/10.1002/cbic.202100240>

## MINIREVIEW

## Core Steps to the Azaphilone Family of Fungal Natural Products

Katherine Williams,<sup>\*,[a]</sup> Claudio Greco,<sup>[b]</sup> Andrew M. Bailey,<sup>[a]</sup> Christine L. Willis.<sup>[c]</sup>

[a] \*Dr. K Williams, Dr. A. M. Bailey  
School of Biological Sciences  
University of Bristol  
Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK  
E-mail: katherine.williams@bristol.ac.uk

[b] Dr. C. Greco  
Department of Molecular Microbiology  
John Innes Centre  
Norwich, NR4 7UH, UK

[c] Prof. C L. Willis  
School of Chemistry  
University of Bristol  
Cantock's Close, Bristol, BS8 1TS, UK

Supporting information for this article is given via a link at the end of the document.

**Abstract:** Azaphilones are a family of polyketide based fungal natural products that exhibit interesting and useful bioactivities. This review explores the literature on various characterised azaphilone biosynthetic pathways, which allows for a proposed consensus scheme for the production of the core azaphilone structure, as well as identifying early diversification steps during azaphilone biosynthesis. A consensus understanding of the core enzymatic steps towards a particular family of fungal natural products can aid in genome mining experiments. Genome mining for novel fungal natural products is a powerful technique for both exploring chemical space and providing new insights into fungal natural product pathways.

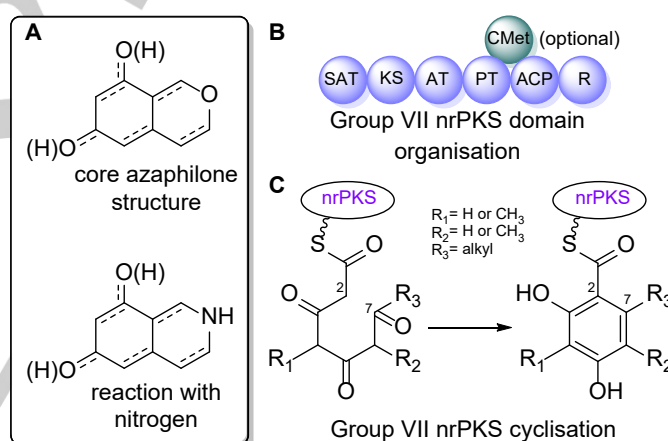
## 1. Introduction.

Azaphilones are a group of structurally related fungal natural products that contain a highly oxygenated bicyclic pyrone-quinone structure, and a chiral quaternary centre. They can react readily with amines to form vinylogous  $\gamma$ -pyridones – hence the name ‘azaphilone’ (Figure 1, A).<sup>[1,2]</sup> Azaphilones exhibit a range of bioactivities including anti-tumour, antifungal and antiviral activity, and are used as food colourants, and as dyes.<sup>[1]</sup>

This review focusses on the biosynthesis of azaphilone compounds and compares common biosynthetic steps between this family of natural products. We examine the biosynthetic pathways to several related compounds (where their biosynthesis informs studies on azaphilones): the important mycotoxin, citrinin **1**,<sup>[3]</sup> and the anti-cancer compound, asperfuranone **2**,<sup>[4]</sup> as well as azaphilone biosynthetic pathways, such as to azanigerone A **3**.<sup>[5]</sup> A consensus understanding of the core reactions that create the azaphilone structure, and the enzymes that catalyse them, can aid the use of genome mining to identify orphan azaphilone and azaphilone-derived biosynthetic gene clusters (BGCs) for further investigation.

The azaphilone and related compounds discussed in this review are synthesised by group VII non-reducing polyketide synthases (nrPKSs) (Figure S1).<sup>[6,7]</sup> These nrPKSs all contain the following domains: SAT (starter unit-ACP transacylase), KS (ketosynthase), AT (acyl transferase), PT (product template), ACP (acyl carrier protein), R (reductase), with some group VII nrPKSs also containing a CMeT (C-methyltransferase) domain (Figure 1, B).

The PT domain of group VII nrPKSs mediate a C-2 to C-7 cyclisation of the initial polyketide chain (Figure 1, C).<sup>[7]</sup> The R domain cleaves the product *via* reductive release leaving an aldehyde group.<sup>[6]</sup>



**Figure 1.** A. Core azaphilone structure, and reaction with nitrogen. B. Domain organisation of group VII nrPKSs C. Cyclisation of the initial polyketide chain.

Dr. Katherine Williams received her PhD from the University of Bristol (UoB) in 2010, before moving to the Bristol Polyketide Group, investigating the biosynthesis of fungal metabolites with interesting bioactive properties. Research posts at the Leibniz Universität Hannover, Germany with Prof. Russell Cox, and at Cardiff University with Prof. Ruedi Allemann followed. Katherine now works back in the UoB, with Dr. Andy Bailey and Prof. Chris Willis, on a project developing a high-throughput heterologous production platform for fungal natural product antibiotic discovery.



## MINIREVIEW

Dr. Claudio Greco received a PhD at the University of Bristol (2017) elucidating the biosynthesis of fungal secondary metabolites under the supervision of Prof. Russell Cox and Prof. Chris Willis. This was followed by a two-year PostDoc at the University of Wisconsin-Madison working with Prof. Nancy Keller, studying secondary metabolism regulation in pathogenic fungi. Claudio is currently working at the John Innes Centre with Prof. Barrie Wilkinson as a BBSRC Discovery Fellow investigating the ecological role of fungal natural products.



Dr. Andy Bailey is a senior lecturer at the University of Bristol, UK. His research interests are based around different aspects of fungal biology and their analysis using molecular genetic approaches. This includes genome mining to explore fungal secondary metabolism, fungi as pathogens of plants and invertebrates and other fungi, plus establishing methods for genetic analysis of basidiomycetes.



Prof. Chris Willis is Professor of Organic Chemistry and Head of Organic and Biological Chemistry at the University of Bristol. Her research focuses on natural product biosynthesis including the application of total synthesis, isotopic labelling, pathway engineering and mechanistic studies to produce biocatalysts and new bioactive molecules. She was the recipient of the Natural Product Chemistry Award of the Royal Society of Chemistry in 2020.

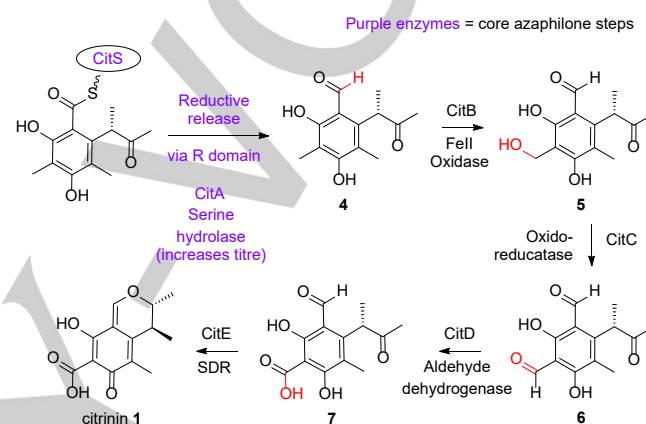


## 2. Elucidation of azaphilone and relevant related biosynthetic pathways.

### 2.1. Citrinin (BGC0001338).

Citrinin **1** is a well-known mycotoxin that is closely related to the azaphilones, but citrinin **1** itself is resistant to amination.<sup>[3]</sup> Early investigations into citrinin **1** biosynthesis used radioisotopes, which confirmed its polyketide origin through the incorporation of <sup>14</sup>C-labelled acetate.<sup>[8,9]</sup> Recent work has defined the molecular steps to citrinin **1** biosynthesis through a series of gene knockout and heterologous expression experiments (Scheme 1).<sup>[3]</sup> Although citrinin **1** bears much structural similarity to azaphilones, consisting of a pyranoquinone bicyclic core, and is in fact often defined as an azaphilone,<sup>[1]</sup> the molecular steps to its structure differentiate at an early stage. A key step, which is also present in some characterised azaphilone biosynthetic pathways, is the

supportive role of the serine hydrolase (CitA in citrinin biosynthesis) to the nrPKS (CitS). He and Cox<sup>[3]</sup> showed that it is required for high titre of the first enzyme free intermediate **4** by both knockout experiments in the producing strain, *Monascus ruber* M7, and heterologous expression experiments. Storm and Townsend<sup>[10]</sup> have further clarified, but not completely elucidated the role of CitA: they show through *in vitro* assays that CitA can remove several enzyme-bound acyl intermediates. They postulated that CitA functions as an *in trans* editing system that can remove acyl-*holo*-ACP species that are not direct intermediates to the programmed PKS product. Knowledge of this initial stage of citrinin biosynthesis informs our understanding of the equivalent steps found in most azaphilone-based pathways. Further steps differ between citrinin and azaphilone biosynthesis.



Scheme 1. Proposed citrinin **1** biosynthetic pathway.

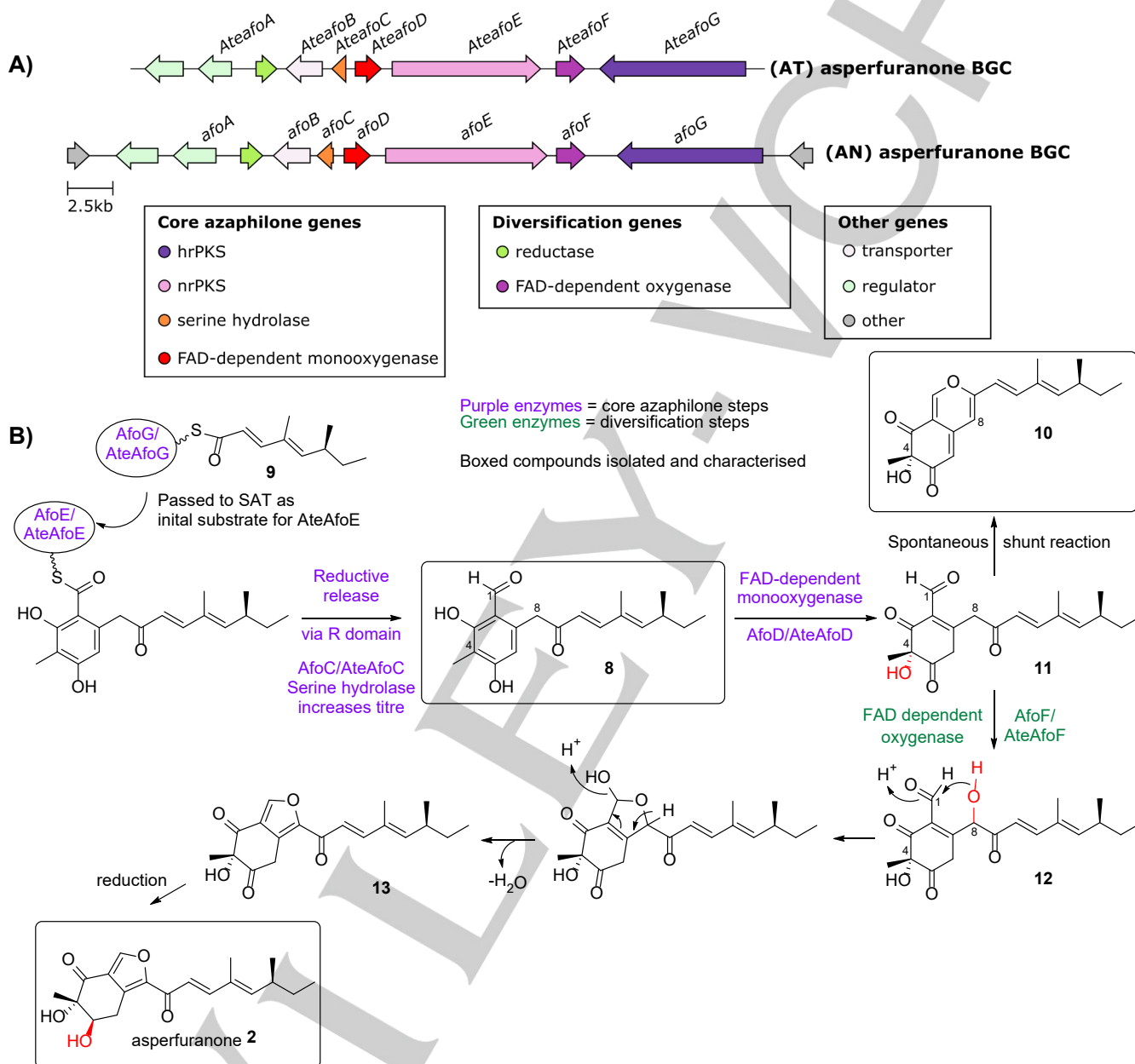
### 2.2. Asperfuranone (BGC000022).

Despite the multitude of known azaphilone compounds isolated from fungi, and extensive labelling studies to investigate azaphilone biosynthesis,<sup>[1,11]</sup> until relatively recently little has been understood about the enzymatic steps, and the encoding genes. In 2009, Wang and co-workers<sup>[12]</sup> identified from the *Aspergillus nidulans* genome the biosynthetic gene cluster (BGC) responsible for the biosynthesis of asperfuranone **2**, a compound which is closely related to the azaphilones, consisting of a bicyclic furan containing core (Figure 2, A). This study provided the first hints towards the molecular pathway of the azaphilones, and showed that several of the enzymes are homologous to those involved in citrinin biosynthesis. Through gene deletion experiments, the authors showed that the serine hydrolase (AfoC) present in the BGC contributes to high titre of the first enzyme free intermediate **8** in a similar manner to citrinin **1** biosynthesis.<sup>[12,3]</sup> The asperfuranone BGC contains both an nrPKS (*afpE*), homologous to *citS*, and a highly reducing PKS (hrPKS - *afpG*). Deletion of either PKS abolishes asperfuranone **2** biosynthesis, leading to the proposal that the hrPKS produces a reduced starter unit **9** that is transferred to the SAT domain of the nrPKS to be further extended.<sup>[12]</sup> Further work in 2013 using an orthologous asperfuranone **2** BGC identified from *Aspergillus terreus* (Figure 2, A), completely reconstituted asperfuranone **2** biosynthesis in a heterologous host and demonstrated the order of each required step (Figure 2, B).<sup>[13]</sup> Co-expression of the hrPKS (*AteAfpG*) and the nrPKS (*AteAfpE*) led to isolation of compound **8** (Figure 2, B),

## MINIREVIEW

demonstrating that both PKSs are required for biosynthesis of the first enzyme free intermediate. The addition of the serine hydrolase (AteAfoC) increased the yield of **8**. This pathway also revealed the essential azaphilone step is catalysed by an FAD-dependent monooxygenase (FMO) with homology to salicylate monooxygenases. In the biosynthesis of asperfuranone, the FMO AteAfoD hydroxylates C-4 of **8**, leading to spontaneous

cyclisation and dehydration to give the shunt azaphilone **10** (Figure 2, B), with the characteristic bicyclic pyrone-quinone structure. The mature compound, asperfuranone **2**, is formed after C-8 hydroxylation of intermediate **11**, by an FAD-dependent oxygenase, AteAfoF, giving **12**. Cyclisation and dehydration gives the five membered furan **13** and reduction leads to asperfuranone **2** (Figure 2, B).<sup>[13]</sup>



**Figure 2.** A) Asperfuranone BGCs from *Aspergillus terreus* (AT) and *Aspergillus nidulans* (AN), core genes required for azaphilone biosynthesis, as well as diversification genes are shown. B) Scheme showing proposed asperfuranone biosynthetic pathway.

### 2.3. Azanigerone A (BGC0001143).

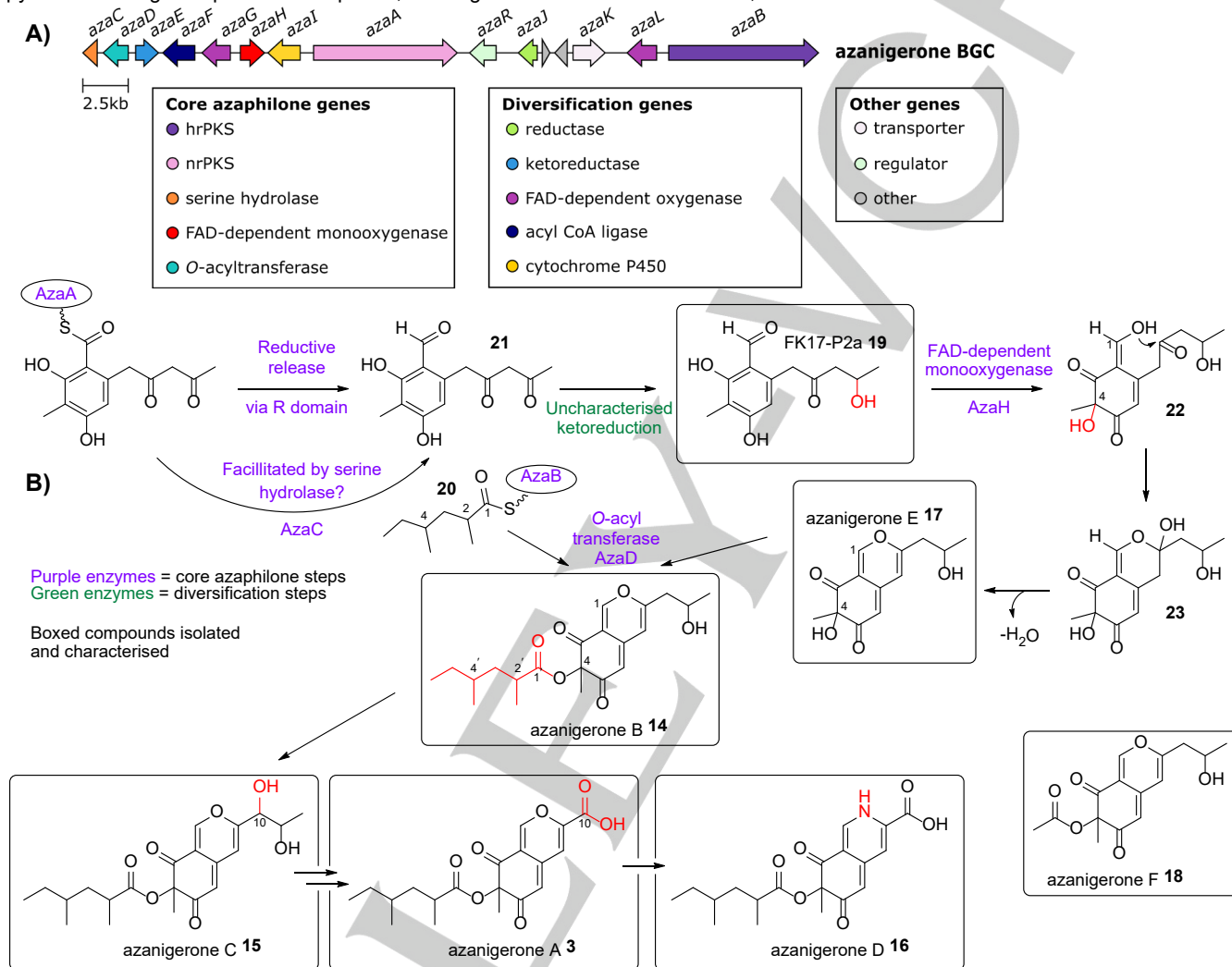
The first BGC encoding azaphilone biosynthesis was reported by Tang and co-workers in 2012.<sup>[5]</sup> A BGC was identified within the *Aspergillus niger* genome which contained both an nrPKS (*azaA*) and a hrPKS (*azaB*) (Figure 3, A), which was minimally expressed

under conditions screened.<sup>[5]</sup> Comparison with the asperfuranone **2** cluster revealed homologues between seven of the catalytic protein sequences from the *aza* BGC, suggesting that the cluster might code for the production of an asperfuranone-like compound. No candidate compounds had previously been isolated from *A. niger*, therefore overexpression of the putative transcription factor present within the BGC was performed. This led to the production

## MINIREVIEW

and isolation of seven compounds, six of which were novel azaphilones, azanigerone A **3** and the azanigerones B-F, **14-18**.<sup>[5]</sup> The seventh compound was the tetrasubstituted benzaldehyde FK17-P2a **19** which has previously been isolated from other fungi.<sup>[14,15]</sup> *In vitro* characterisation of the FAD-dependent monooxygenase (AzaH, a homologue of AteAfoD) demonstrated that is solely responsible for the flavin-dependent conversion of the benzaldehyde intermediate FK17-P2a **19** into the bicyclic, pyran-containing azaphilone compound, azanigerone E **17**

(Figure 3, B). Deletion of the hrPKS (*azaB*), in contrast to the asperfuranone pathway, led to the accumulation of azanigerone E **17**. This suggested that AzaB is only responsible for the biosynthesis of the 2,4-dimethylhexanoyl chain **20**, which is proposed to be the substrate for the *O*-acyltransferase AzaD to esterify the C-4 hydroxyl group of azanigerone E **17** to produce azanigerone B **14**. Additional predictions were made about the pathway, based on a time course experiment and isolation of the metabolites, but not further verified.<sup>[5]</sup>



**Figure 3.** A) Azanigerone BGC, core genes required for azaphilone biosynthesis, as well as diversification genes are shown. B) Scheme showing proposed azanigerone biosynthetic pathway.

## 2.4. Chaetoviridins and chaetomugilins (BGC0001405).

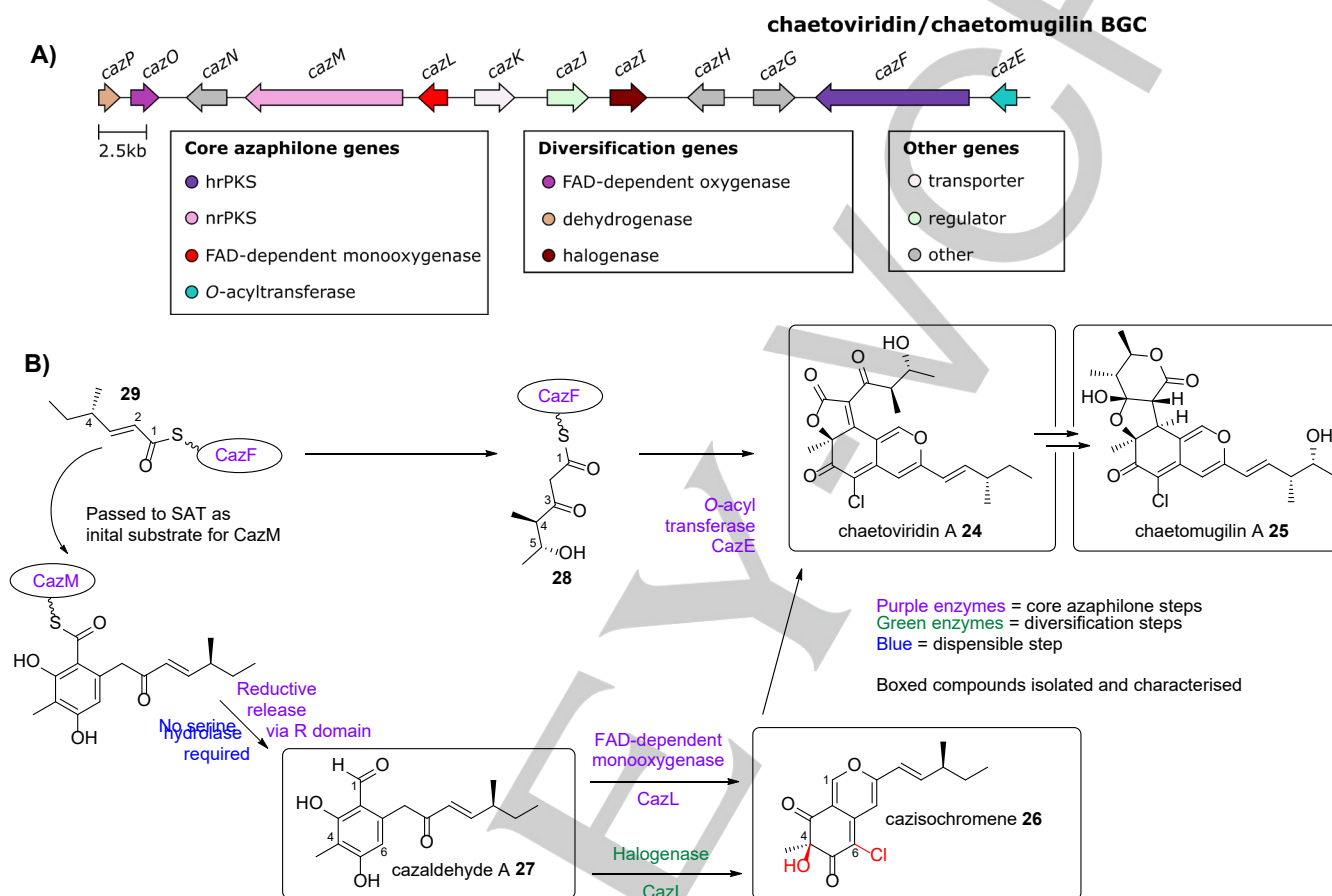
Characterisation of a second azaphilone pathway was reported in 2012 by Tang and co-workers,<sup>[16]</sup> following studies on the biosynthesis of chlorinated azaphilone compounds produced by *Chaetomium globosum*, for example chaetoviridin A **24**<sup>[17]</sup> and chaetomugilin A **25**.<sup>[18]</sup> The BGC that codes for production of these compounds was identified from the genome by searching for clustered homologues of the nrPKSs and hrPKSs present in the asperfuranone and azanigerone BGCs (Figure 4, A). Knock out of either the nrPKS (*cazM*) or the hrPKS (*cazF*) abolished production of any azaphilone compounds, with no other

intermediates accumulating. Knock-out of the *O*-acyltransferase, *cazE* (which has a homologue in the azanigerone A **3** BGC, *azaD*) accumulated cazisochromene **26** (Figure 4, B), which is structurally related to azanigerone E **17** (an intermediate in the azanigerone A **3** pathway). Cazisochromene **26** is likely to be the product of C-4 hydroxylation by *CazL* and C-6 chlorination by *CazI* of cazaldehyde A **27**. Timing of chlorination versus hydroxylation is unclear. Interestingly, no serine hydrolase is present in this BGC, suggesting that its activity is not always essential in azaphilone biosynthesis. *In vitro* assays using *CazF* (hrPKS), *CazE* (*O*-acyltransferase), and cazisochromene **26**, resulted in the production of chaetoviridin A **24**. This is in accord with the hrPKS giving the 5-hydroxy-4-methyl-3-oxohexanoyltriketide **28**

## MINIREVIEW

that is required to produce the lactone ring fused to the isochromenone core of chaetoviridin A **24**. The lack of any azaphilone compounds produced by the  $\Delta\text{cazF}$  strain suggested that the hrPKS is also implicated in the production of the 4-methylhex-2-enoyltriketide starter unit **29** required by the nrPKS (CazM).<sup>[16]</sup> This was confirmed in a later study,<sup>[19]</sup> whereby *in vitro* assays of CazF and CazM produced cazaldehyde A **27**, the first

enzyme free intermediate that undergoes hydroxylation and chlorination before cyclisation and dehydration forms the azaphilone core. Experiments showed that the SAT domain only recognised the more reduced triketide.<sup>[19]</sup> This is an interesting example of an hrPKS producing two differing substrates required for the biosynthesis of the mature compound.



**Figure 4.** A) Chaetoviridins and chaetomugilins BGC, core genes required for azaphilone biosynthesis, as well as diversification genes are shown. B) Scheme showing proposed chaetoviridin A biosynthetic pathway. Carbons are numbered to correspond to azanigerone A **3**.

#### 2.4. *Monascus* spp. azaphilones (BGC0000027).

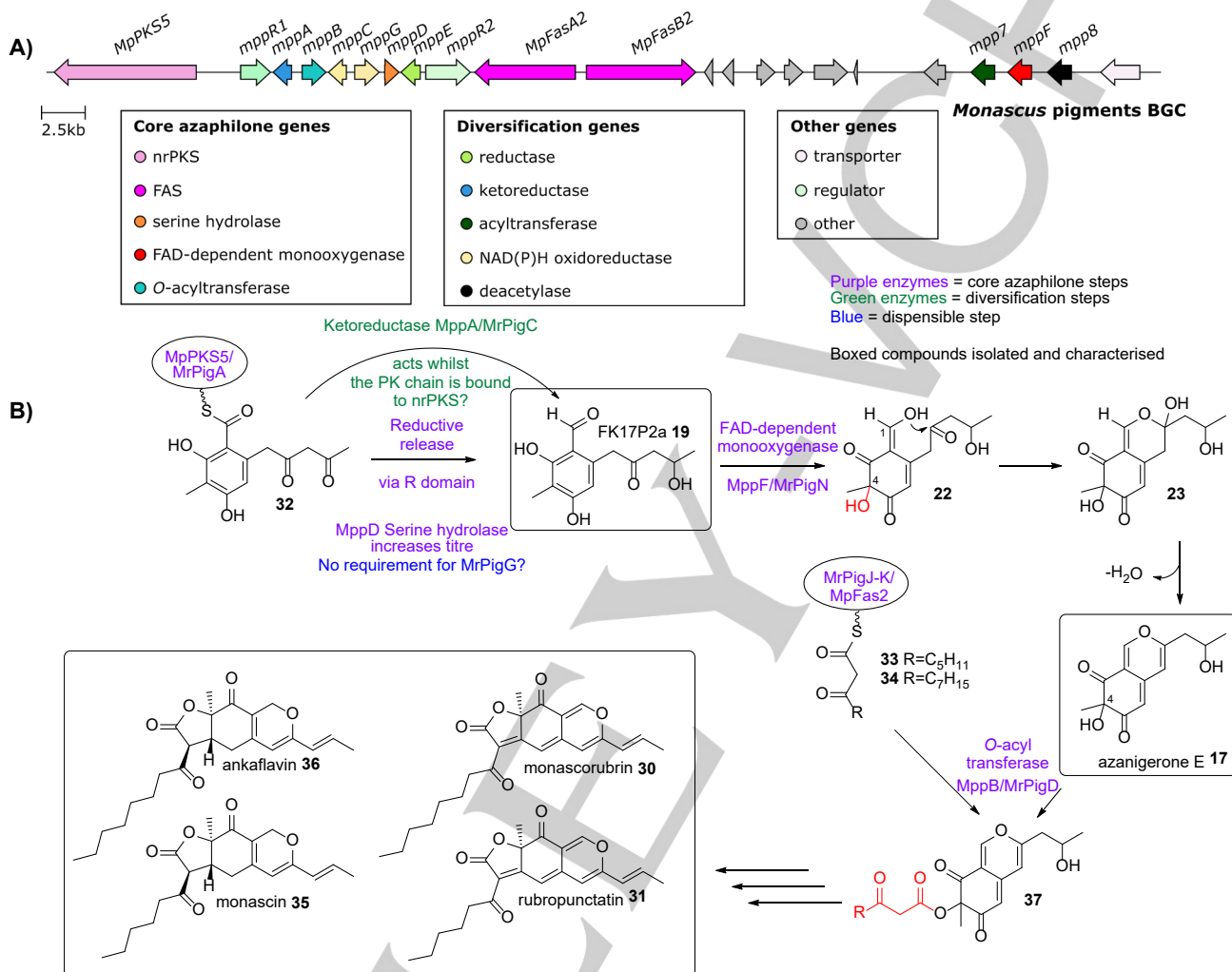
The BGC for the well-known food colourant azaphilones isolated from *Monascus* spp. such as monascorubrin **30** and rubropunctatin **31** was reported in 2013 by Kwon and co-workers (Figure 5, A).<sup>[20]</sup> This was achieved by random T-DNA mutagenesis of *Monascus purpureus* and screening for pigment loss. At the time, the genome of *M. purpureus* had not been sequenced, therefore the available *Monascus pilosus* genome was used to align the sequence of the T-DNA insertion region to identify a putatively orthologous BGC. Once identified, the nrPKS present in the cluster (*MpPKS5*), which is homologous to other known azaphilone nrPKSs, was targeted for gene deletion in *M. purpureus*, to create strain  $\Delta\text{MpPKS5}$ . This strain no longer produced any azaphilones, demonstrating that *MpPKS5* is essential to azaphilone biosynthesis in *M. purpureus*. Kwon and co-workers<sup>[21]</sup> later showed that the serine hydrolase (*MppD*)

present in the BGC contributes to high titre of the first enzyme free intermediate **19** in a similar manner to citrinin **1** and asperuranone **2** biosynthesis.<sup>[21,3,12]</sup> Interestingly, an orthologous cluster from *Monascus ruber* studied by Chen and co-workers<sup>[22]</sup> did not appear to possess the same reliance on the serine hydrolase for high titre. Co-expression of the nrPKS (*MrPigA*) with the serine hydrolase (*MrPigG*) in the heterologous host *Saccharomyces cerevisiae* did not increase yield compared to expression of *MrPigA* alone.<sup>[22]</sup> Studies by both groups show that the orthologs *MppA* and *MrPigC* catalysed the ketoreduction of the highly reactive benzaldehyde **32** to give **19**.<sup>[23,22]</sup> There is no homologue to *MppA*/*MrPigC* in the azanigerone A **3** cluster, although the same ketoreduction does occur, and Tang and co-workers speculated that *AzaE* is instead responsible.<sup>[5]</sup> The FAD-dependent monooxygenase orthologues *MppF*/*MrPigN* catalyse the C-4 ring hydroxylation of FK17-P2a **19** to form azanigerone E **17** (Figure 5, B).<sup>[22,23]</sup> In these BGCs a fatty acid synthase dimer,

## MINIREVIEW

MpFas2/MrPigJ-K catalyses the biosynthesis of the fatty acid moieties **33** and **34**, which are present in the *Monascus* spp. azaphilones, unlike the hrPKSs that are present in the azanigerone A **3** or the chaetoviridin/chaetomugilins BGCs.<sup>[24,22]</sup> An O-acyltransferase, MrPigD, homologous to those found in the azanigerone A **3** and the chaetoviridin/chaetomugilins BGCs, transfers the fatty acid moieties to the C-4 hydroxyl group. This

was demonstrated by the creation of a  $\Delta$ MrPigD strain which accumulates azanigerone E **17**.<sup>[22]</sup> The O-acyltransferase homologue MppA from the *Monascus purpureus* BGC is proposed to perform the same reaction.<sup>[24]</sup> Further steps lead to generation of monascin **35**, rubropunctatin **31**, ankaflavin **36** and monascorubrin **30**.<sup>[22,25,26,27]</sup>



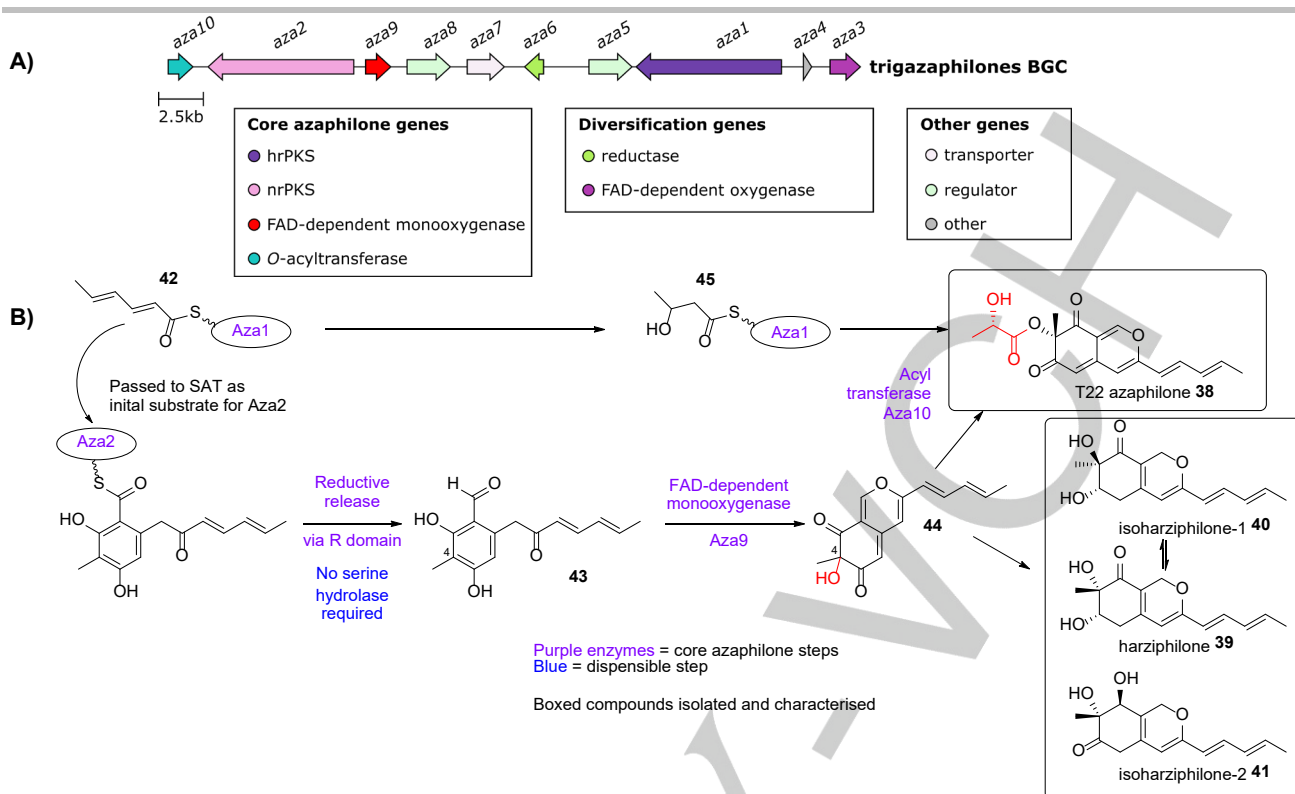
**Figure 5.** A) *Monascus* pigments BGC, core genes required for azaphilone biosynthesis, as well as diversification genes are shown. B) Scheme showing proposed *Monascus* spp. azaphilone biosynthetic pathway.

## 2.5. Trigazaphilones (LVVK01000021.1, 680,000–727,296).

Recently, a BGC identified from the mycoparasitic fungus *Trichoderma guizhouense* was shown to encode the biosynthesis of azaphilones. Zhang and co-workers<sup>[28]</sup> noticed the production of a 'light yellow pigment' when *T. guizhouense* interacts with and inhibits the growth of *Fusarium odoratissimum*. Through analysis of a transcriptome profile of this interaction,<sup>[29]</sup> upregulated BGCs were identified, one of which was shown to be responsible for the biosynthesis of the 'light yellow pigment', through deletion of the two PKS genes encoded within the BGC (*aza1* and *aza2*, Figure 6, A). The compounds were isolated and characterised as T22

azaphilone **38**, harziphilone **39**, isoharziphilone-1 **40** and isoharziphilone-2 **41**, which collectively have been named the trigazaphilones (Figure 6, B). Deletion of the O-acyltransferase encoded with the BGC (*aza10*) abolished T22 azaphilone **38** biosynthesis, with the deletion strain accumulating harziphilone **39** and its isomers **40** and **41**. The authors proposed that the biosynthesis proceeds in a similar manner to that of the chaetoviridins and chaetomugilins (Figure 6, B). Interestingly, this BGC represents another azaphilone biosynthetic pathway which has no serine hydrolase, supporting the dispensible nature of this enzyme.

## MINIREVIEW

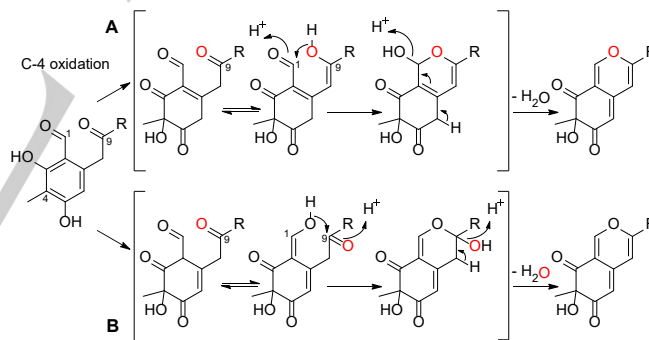


**Figure 6.** A) Trigazaphilones BGC, core genes required for azaphilone biosynthesis, as well as diversification genes are shown. B) Scheme showing proposed trigazaphilones biosynthetic pathway.

### 3. Consensus pathway.

The current consensus (Scheme 2) for the core steps required for azaphilone biosynthesis are exemplified by the azanigerone **A 3** pathway. In fact, azanigerone **E 17**, an intermediate in azanigerone **A 3** biosynthesis (Figure 3, B) is also an intermediate in *Monascus* spp. pigment biosynthesis (Figure 5, B), furthermore, the structurally related cazisochromene **26**, is an intermediate in the biosynthesis of chaetoviridins and chaetomugilins (Figure 4, B), as well as a similar analogue (**10**) being produced as a shunt compound during asperfuranone **2** biosynthesis (Figure 2, B). Azanigerone **E 17** (or analogues) require an nrPKS to biosynthesise a benzaldehyde intermediate, the titre of which may be increased by the serine hydrolase (although it is not always necessary). The nrPKS may require an hrPKS to produce a specific reduced starter unit for further extension.

The FAD-dependent monooxygenase catalyses C-4 hydroxylation, which is the key step in azaphilone biosynthesis, converting the benzaldehyde into the bicyclic, pyran-containing core of azaphilones. It is interesting to note that two mechanisms have been suggested for the biosynthesis of this core structure. Wang and co-workers<sup>[13]</sup> proposed that the C-1 aldehyde group acts as an electrophile, whilst Tang and co-workers<sup>[5]</sup> proposed that the same group acts as a nucleophile. Labelling studies could distinguish the correct mechanism (Figure 7).



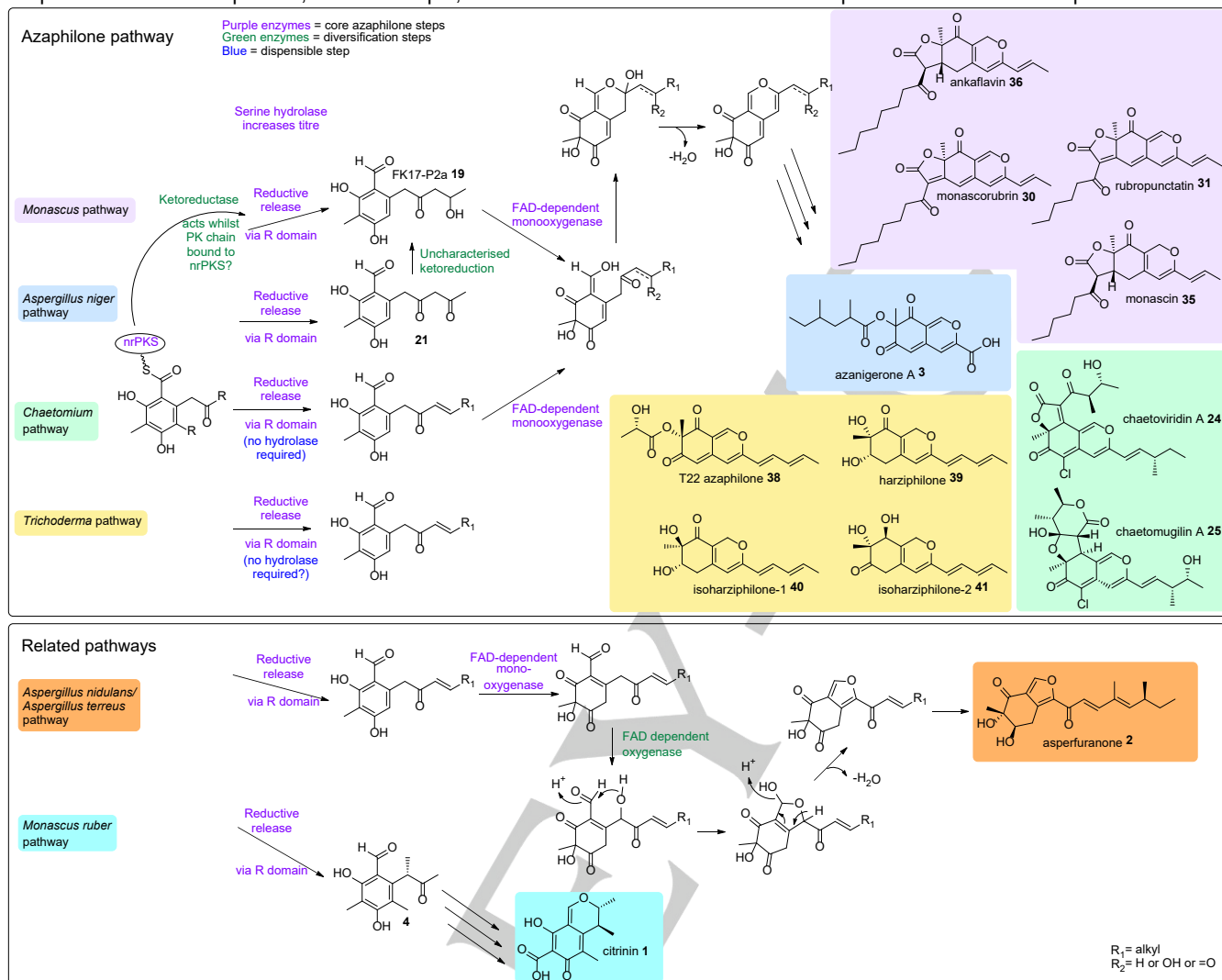
**Figure 7.** Differing azaphilone bicyclic ring formation mechanisms. Mechanism A was proposed by Wang and co-workers,<sup>[13]</sup> mechanism B was proposed by Tang and co-workers.<sup>[5]</sup> A labelling study (labelled oxygen depicted in red) would distinguish the correct mechanism.

The C-4 hydroxyl group introduced by the FAD-dependent monooxygenase is then a target for a further common step in azaphilone biosynthesis, acylation of the oxygen by an O-acyltransferase. The acyl moiety can be biosynthesised by either an hrPKS (which can be the same hrPKS that produces the starter unit, as seen in Figure 4, B), or an FAS, as seen *Monascus* spp. pigment biosynthesis (Figure 5, B). Diversification of azaphilone structures can occur in these early stages, for example, the requirement for a starter unit for the nrPKS, which can vary in length and reduction, or the halogenation that occurs in chaetoviridin and chaetomugilin pathway (Figure 4, B). Further diversification can also occur after the transfer of the acyl moiety, to produce the range of azaphilone structures, some of which are



## MINIREVIEW

illustrated in Scheme 2, as well as other more complex azaphilone-based compounds, for example, the dimeric azaphilones.<sup>[30]</sup> A recent review by Franck and co-workers discusses various azaphilone diversification steps.<sup>[31]</sup>

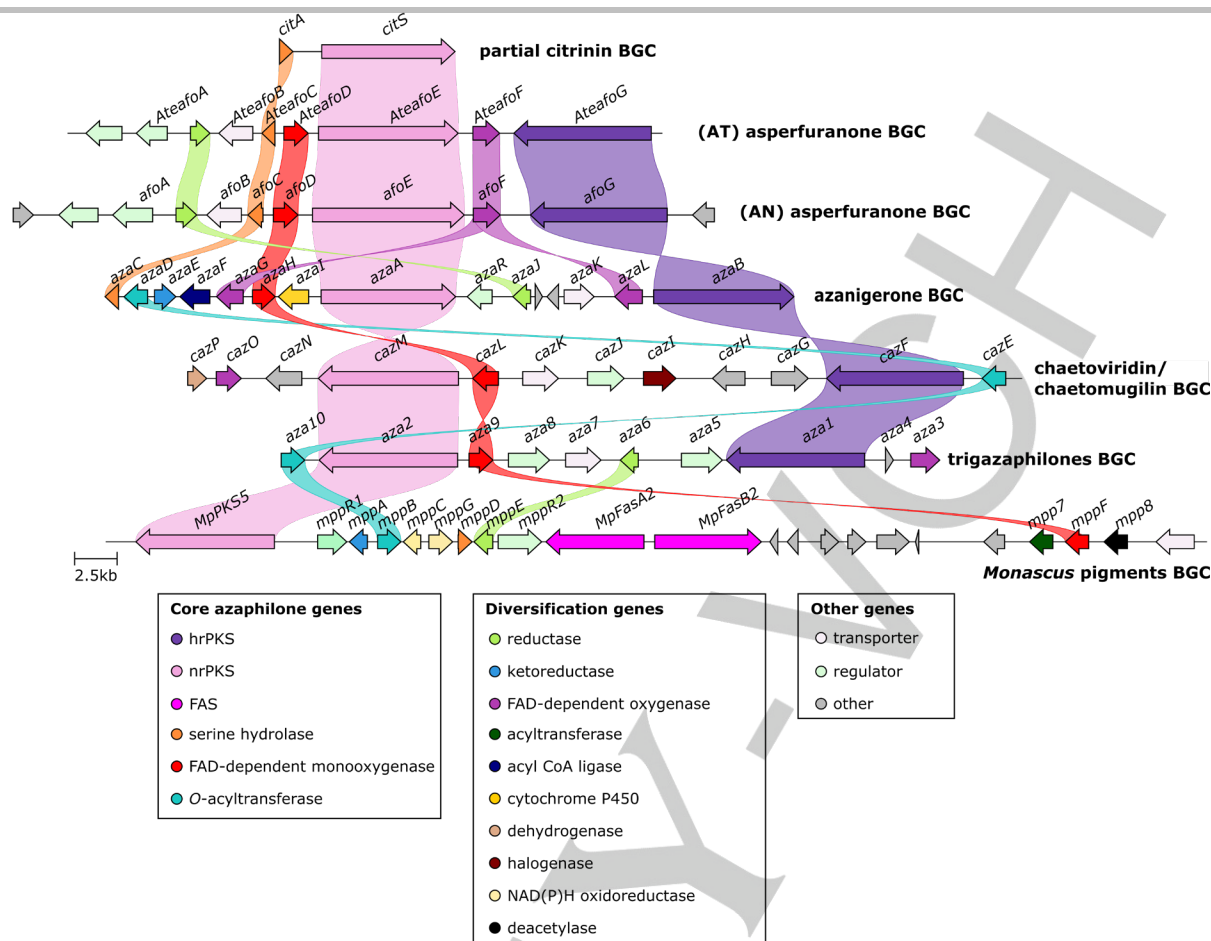


**Scheme 2.** A comparison of azaphilone and several related pathways.

A clinker<sup>[32]</sup> comparative analysis of the BGCs discussed in sections 2.1-2.5 (Figure 8) shows the similarities and differences between these clusters. It is clear that each gene is not always necessary for azaphilone biosynthesis, but we hypothesise that their clustered presence within a fungal genome would be highly indicative of a BGC that codes for the production of an azaphilone

compound. Core protein sequences that can be used for identifying further azaphilone BGCs from genome sequences are AzaA (nrPKS), AzaB (hrPKS), AzaC - serine hydrolase (SH), AzaH - FAD-dependent monooxygenase (FMO) and AzaD - O-acyltransferase (OAT) (Table 1).

## MINIREVIEW



**Figure 8.** clinker<sup>[32]</sup> comparison of the azaphilone and related BGCs described above. For the citrinin BGC, only the relevant azaphilone related genes are shown. The *Monascus pigments Monascus ruber* BGC genomic sequence was not publicly available for comparison. The citrinin sequences are from *Monascus ruber* (MR). The asperfuranone sequences are from either *Aspergillus nidulans* (AN) or *Aspergillus terreus* (AT).

**Table 1:** Percentage identity between protein sequences from core azaphilone and related pathways. nrPKS = non reducing polyketide synthase, hrPKS = highly reducing polyketide synthase, SH = serine hydrolase, FMO = FAD-dependent monooxygenase, OAT = O-acyltransferase, FAS- $\alpha/\beta$  = fatty acid synthase subunit. Accession numbers for each protein are shown underneath the protein name. The citrinin sequences are from *Monascus ruber* (MR). The asperfuranone sequences are from either *Aspergillus nidulans* (AN) or *Aspergillus terreus* (AT). The *Monascus pigments* sequences are from either *Monascus pilosus* (MPi) or from *Monascus ruber* (MR). \*MppD was originally annotated incorrectly, and the sequence was later revised, but the revised sequence is not publicly available, therefore the sequence was manually derived via homology searches (see Figure S2).<sup>[25]</sup>

Function	Azanigerone	% identity compared to azanigerone core protein sequences						
		Citrinin (MR)	Asperfuranone (AN)	Asperfuranone (AT)	Monascus pigments (MPi)	Monascus pigments (MR)	Chaetoviridin	Trigazaphilones
nrPKS	<b>AzaA</b> G3XMC4.1	CitS (47.87) A0A161CEU9.1	AfoE (45.66) Q5BEJ6.1	AteAfoE (46.39) Q0CF73.1	MpPKS5 (58.12) AGN71604.1	MrPigA (58.23) ALN44200.1	CazM (46.05) AKA40070.1	Aza2 (46.64) OPB37950.1
hrPKS	<b>AzaB</b> G3XMD1.1	n/a	AfoG (47.14) Q5BEJ4.1	AteAfoG (47.54) Q0CF75.1	n/a	n/a	CazF (47.95) EAQ86385.1	Aza1 (44.82) OPB37944.1
SH	<b>AzaC</b> G3XMB7.1	CitA (47.28) A0A161CKG1.1	AfoC (48.16) Q5BEJ8.1	AteAfoC (51.26) Q0CF71.1	MppD* (57.32)	MrPigG (58.54) AP273942.1	n/a	n/a
FMO	<b>AzaH</b> G3XMC2.1	n/a	AfoD (41.05) Q5BEJ7.1	AteAfoD (41.63) Q0CF72.1	MppF (58.27) AGN71623.1	MrPigN (59.23) ALT31754.1	CazL (42.08) EAQ86391.1	Aza9 (37.68) OPB37949.1
OAT	<b>AzaD</b> G3XMB8.1	n/a	n/a	n/a	MppB (49.22) AGN71607.1	MrPigD (49.22) AGI63864.1	CazE (33.79) EAQ86384.1	Aza10 (33.04) OPB37951.1
FAS- $\alpha$	<b>n/a</b>	n/a	n/a	n/a	MpFasA2 AGN71613.1	MrPigJ AGL44429.1	n/a	n/a
FAS- $\beta$	<b>n/a</b>	n/a	n/a	n/a	MpFasB2 AGN71614.1	MrPigK AGL44430.1	n/a	n/a

## MINIREVIEW

To test this theory, we have conducted a representative cblaster<sup>[33]</sup> search for co-localised genes from annotated genomes on the NCBI database, using the above five sequences as queries, with AzaA, AzaD and AzaH set as required (Figure S3). This search detected 193 BGCs with azaphilone producing potential, 8 of which have been either definitively<sup>[5,34-36,28]</sup> or putatively<sup>[30,37]</sup> linked to the biosynthesis of azaphilone derived compounds. This leaves 185 orphan BGCs that do not appear yet to be linked to specific compound biosynthesis, however many of the species have known azaphilone producers within their genera.<sup>[11,31,2]</sup> Orphan BGCs are candidates for further investigation into azaphilone biosynthesis and production. Other cblaster<sup>[33]</sup> searches could be attempted, for example, without AzaD set as required (as the O-acylation step is not necessary for the formation of the bicyclic pyrone-quinone azaphilone core), or alternatively using the MpFas2/MrPigJ-K fatty acid synthase dimer sequences instead of AzaB, or using diversifying enzymes sequences, for example CazI, the halogenase from the chaetoviridins and chaetomugilins BGC.

#### 4. Summary and Outlook.

Fungal genome mining is an exciting method to help realise the full natural product potential revealed within fungal genomes. Many tools exist to facilitate fungal genome screening, two of the most popular are fungiSMASH<sup>[38]</sup> and SMURF,<sup>[39]</sup> both of which use profile Hidden Markov Model (pHMM) based algorithms to identify BGCs, along with other complementary algorithms such as CASSIS (uses co-regulatory motifs),<sup>[40]</sup> CO-OCCUR (shared syntenic relationships among genes)<sup>[41]</sup> and FunOrder<sup>[42]</sup> (co-evolutionary linked genes). An understanding of the common biosynthetic steps towards a particular family of compounds can help identify specific BGCs amongst the wealth of data that the number of sequenced fungal genomes provides. Co-localisation of homologues to specific sequences from the azanigerone BGC, (AzaA (nrPKS), AzaB (hrPKS), AzaC (serine hydrolase), AzaH (FAD-dependent monooxygenase) and AzaD, (O-acyltransferase) see Table 1) is highly likely to indicate the presence of an azaphilone BGC within a fungal genome. Many of the species identified by the cblaster<sup>[33]</sup> analysis shown in Figure S3 are not specifically known to produce azaphilones.<sup>[2,11,31]</sup> Investigation of orphan azaphilone BGCs, especially in species where azaphilones have not previously been detected (and therefore might indicate unexplored chemical space) could be a fruitful avenue for the discovery and isolation of novel azaphilones.

#### Acknowledgements.

This work was supported by the MRC (MR/N029909/1 - KW), and the BBSRC (BB/N005723/1 - CG). KW thanks Dr. Kate de Mattos-Shiple for useful discussions on this work. The authors would like to thank the reviewers for their useful and insightful comments.

**Keywords:** natural products • biosynthesis • genome mining • azaphilone • pathway elucidation

[1] J. M. Gao, S. X. Yang, J. C. Qin, *Chem. Rev.* **2013**, *113*, 4755-4811.

[2] N. Osmanova, W. Schultze, N. Ayoub, *Phytochem. Rev.* **2010**, *9*, 315-342.

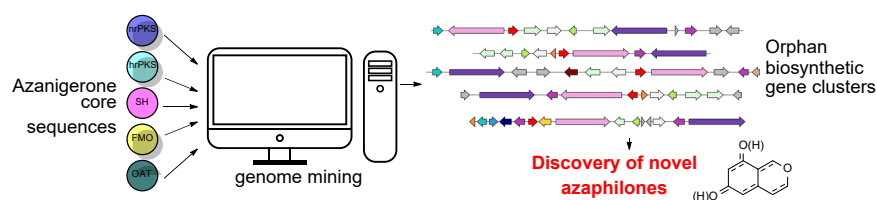
- [3] Y. He, R. J. Cox, *Chem. Sci.* **2016**, *7*, 2119-2127.
- [4] C. C. C. Wang, Y. M. Chiang, M. B. Praseuth, P. L. Kuo, H. L. Liang, Y. L. Hsu, *Basic Clin. Pharmacol. Toxicol.* **2010**, *107*, 583-589.
- [5] A. O. Zabala, W. Xu, Y. H. Chooi, Y. Tang, *Chem. Biol.* **2012**, *19*, 1049-1059.
- [6] 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, C. C. C. Wang, *J. Am. Chem. Soc.* **2012**, *134*, 8212-8221.
- [7] L. Liu, Z. Zhang, C.-L. Shao, J.-L. Wang, H. Bai, C.-Y. Wang, *Sci. Rep.* **2015**, *5*, 10463.
- [8] E. Schwenk, G. J. Alexander, A. M. Gold, D. F. Stevens, *J. Biol. Chem.* **1958**, *233*, 1211-1213.
- [9] A. J. Birch, P. Fitton, E. Pride, A. J. Ryan, H. Smith, W. B. Whalley, *J. Chem. Soc.* **1958**, 4576-4581.
- [10] P. A. Storm, C. A. Townsend, *Chem. Commun.* **2018**, *54*, 50-53.
- [11] C. M. Chen, H. M. Tao, W. H. Chen, B. Yang, X. F. Zhou, X. W. Luo, Y. H. Liu, *RSC Adv.* **2020**, *10*, 10197-10220.
- [12] Y. M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley, C. C. C. Wang, *J. Am. Chem. Soc.* **2009**, *131*, 2965-2970.
- [13] Y. M. Chiang, C. E. Oakley, M. Ahuja, R. Entwistle, A. Schultz, S. L. Chang, C. T. Sung, C. C. C. Wang, B. R. Oakley, *J. Am. Chem. Soc.* **2013**, *135*, 7720-7731.
- [14] Y. Yamaguchi, R. Masuma, Y.-P. Kim, R. Uchida, H. Tomoda, S. Omura, *Mycoscience* **2004**, *45*, 9-16.
- [15] T. Arai, H. Sano, (Kaiyo Bio Technol Kenkyusho) JPH06329576A **1994**.
- [16] J. M. Winter, M. Sato, S. Sugimoto, G. Chiou, N. K. Garg, Y. Tang, K. Watanabe, *J. Am. Chem. Soc.* **2012**, *134*, 17900-17903.
- [17] M. Makrerougras, R. Coffinier, S. Oger, A. Chevalier, C. Sabot, X. Franck, *Org. Lett.* **2017**, *19*, 4146-4149.
- [18] T. Yamada, M. Doi, H. Shigeta, Y. Muroga, S. Hosoe, A. Numata, R. Tanaka, *Tetrahedron Lett.* **2008**, *49*, 4192-4195.
- [19] J. M. Winter, D. Cascio, D. Dietrich, M. Sato, K. Watanabe, M. R. Sawaya, J. C. Vederas, Y. Tang, *J. Am. Chem. Soc.* **2015**, *137*, 9885-9893.
- [20] B. Balakrishnan, S. Karki, S. H. Chiu, H. J. Kim, J. W. Suh, B. Nam, Y. M. Yoon, C. C. Chen, H. J. Kwon, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 6337-6345.
- [21] B. Balakrishnan, R. Chandran, S. H. Park, H. J. Kwon, *J. Microbiol. Biotechnol.* **2015**, *25*, 1648-1652.
- [22] W. P. Chen, R. Chen, Q. P. Liu, Y. He, K. He, X. L. Ding, L. J. Kang, X. X. Guo, N. N. Xie, Y. X. Zhou, Y. Y. Lu, R. J. Cox, I. Molnar, M. Li, Y. C. Shao, F. S. Chen, *Chem. Sci.* **2017**, *8*, 4917-4925.
- [23] B. Bijinu, J. W. Suh, S. H. Park, H. J. Kwon, *RSC Adv.* **2014**, *4*, 59405-59408.
- [24] B. Balakrishnan, H. J. Kim, J. W. Suh, C. C. Chen, K. H. Liu, S. H. Park, H. J. Kwon, *J. Korean Soc. Appl. Biol. Chem.* **2014**, *57*, 191-196.
- [25] B. Balakrishnan, C. C. Chen, T. M. Pan, H. J. Kwon, *Tetrahedron Lett.* **2014**, *55*, 1640-1643.
- [26] B. Balakrishnan, S. H. Park, H. J. Kwon, *Appl. Biol. Chem.* **2017**, *60*, 437-446.
- [27] B. Balakrishnan, S. H. Park, H. J. Kwon, *Biotechnol. Lett.* **2017**, *39*, 163-169.
- [28] G. Pang, T. T. Sun, Z. Z. Yu, T. Yuan, W. Liu, H. Zhu, Q. Gao, D. Q. Yang, C. P. Kubicek, J. Zhang, Q. R. Shen, *Environ. Microbiol.* **2020**, *22*, 4808-4824.
- [29] J. Zhang, Y. Z. Miao, M. J. Rahimi, H. Zhu, A. Steindorff, S. Schiessler, F. Cai, G. Pang, K. Chenthamara, Y. Xu, C. P. Kubicek, Q. R. Shen, I. S. Druzhinina, *Environ. Microbiol.* **2019**, *21*, 2644-2658.
- [30] K. Becker, S. Pflutze, E. Kuhnert, R. J. Cox, M. Stadler, F. Surup, *Chem. - Eur. J.* **2021**, *27*, 1438-1450.
- [31] C. Pavesi, V. Flon, S. Mann, S. Leleu, S. Prado, X. Franck, *Nat. Prod. Rep.* **2021**. DOI: 10.1039/d0np00080a
- [32] C. L. M. Gilchrist, Y.-H. Chooi, *Bioinformatics* **2021**, 1-3.
- [33] C. L. M. Gilchrist, T. J. Booth, Y.-H. Chooi, **2020**, bioRxiv preprint DOI: 10.1101/2020.11.08.370601.
- [34] B. Wang, X. Li, D. Yu, X. Chen, J. Tabudravu, H. Deng, L. Pan, *Microbiol. Res.* **2018**, *217*, 101-107.
- [35] P. C. Y. Woo, C.-W. Lam, E. W. T. Tam, C. K. F. Leung, S. S. Y. Wong, S. K. P. Lau, K.-Y. Yuen, *PLoS Neglected Trop. Dis.* **2012**, *6*, e1871.

## MINIREVIEW

- [36] T. Nakazawa, K. i. Ishiuchi, M. Sato, Y. Tsunematsu, S. Sugimoto, Y. Gotanda, H. Noguchi, K. Hotta, K. Watanabe, *J. Am. Chem. Soc.* **2013**, *135*, 13446-13455.
- [37] G. J. Kramer, S. Pimentel-Elardo, J. R. Nodwell, *ChemBioChem* **2020**, *21*, 2116-2120.
- [38] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81-W87.
- [39] N. Khaldi, F. T. Seifuddin, G. Turner, D. Haft, W. C. Nierman, K. H. Wolfe, N. D. Fedorova, *Fungal Genet. Biol.* **2010**, *47*, 736-741.
- [40] T. Wolf, V. Shelest, N. Nath, E. Shelest, *Bioinformatics* **2016**, *32*, 1138-1143.
- [41] E. Gluck-Thaler, S. Haridas, M. Binder, I. V. Grigoriev, P. W. Crous, J. W. Spatafora, K. Bushley, J. C. Slot, *Mol. Biol. Evol.* **2020**, *37*, 2838-2856.
- [42] G. A. Vignolle, D. Schaffer, R. L. Mach, A. R. Mach-Aigner, C. Derntl, **2021**, bioRxiv preprint DOI: 10.1101/2021.01.29.428829.

## MINIREVIEW

## Entry for the Table of Contents



Azaphilones are a polyketide-based family of fungal natural products with a highly oxygenated bicyclic pyrone-quinone structure, many of which have useful and interesting bioactivities. This review reveals the common steps in azaphilone biosynthetic pathways to aid the use of genome mining to identify orphan azaphilone biosynthetic gene clusters.

Institute and/or researcher Twitter usernames: @Willis\_Group, @BristolBioSci, @BristolUni, @kw2990, @ClaudioGreko