Genome mining of a fungal endophyte of *Taxus yunnanensis* (Chinese yew) leads to the discovery of a novel azaphilone polyketide, lijiquinone

Jesse W. Cain, Kristin I. Miller, John A. Kalaitzis, Rocky Chau, and Brett A. Neilan

1School of Biotechnology and Biomolecular Sciences, UNSW Australia, Sydney, NSW 2052, Australia.
2School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW2308, Australia.

Introduction

Traditional Chinese medicine (TCM) plants have been used for thousands of years to treat a wide range of ailments and diseases, and their use in modern medicine is based on an extensive body of inherited and documented traditional knowledge, which is still practiced today (Qin and Xu, 1998; Wang et al., 2012; Lee et al., 2013). TCM plants have provided the basis for many modern day pharmaceuticals, including the anticancer drugs paclitaxel and camptothecin, isolated from *Taxus* (yew) species and *Camptotheca acuminata* ‘happy tree’, respectively, the antimalarial drug artemisinin, isolated from *Atemnesia annua* (sweet wormwood), the antiviral drug podophyllotoxin, isolated from *Podophyllum peltatum* (American mayapple) and the decongestant pseudoephedrine, isolated from *Ephedra sinica* (ma huang) (Klayman et al., 1984; Wall and Wani, 1996; White et al., 1997; Eyberger et al., 2006; Lee et al., 2013). As a result of such discoveries, TCM plants continue to be intensively investigated as sources of novel bioactive natural products (Adams and Lien, 2013).

Plants are home to diverse communities of microorganisms that are dominated by fungi and bacteria (Miller et al., 2012b). These plant-associated microbes, sometimes termed endophytes, are themselves a rich source of structurally novel bioactive natural products (Gunatilaka, 2006), predominantly polyketides and non-ribosomal peptides (Fischbach and Walsh, 2006). These are structure classes that often give rise to clinically relevant pharmaceuticals (Clardy et al., 2006; Newman and Cragg, 2012; Butler et al., 2014). Such is the current understanding of microbial polyketide and non-ribosomal peptide biosynthesis, many researchers now hypothesize that some natural products originally thought to be plant products are likely biosynthesized by their endophytic microbes (Stierle et al., 1994; Li et al., 1996; Aly et al., 2010; Aly et al., 2013).

The cultivation and extraction of bioactive compounds from microbial primary producers (i.e. endophytes) represents an efficient and sustainable alternative to harvesting biomass from traditional medicinal plants, which are often slow-growing and rare or endangered, as is the case for most *Taxus* species. Endophytes may be screened for bioactive compounds, or the potential to produce such compounds, using bioassays, analytical

### Reference

Adams, P. D. and Lien, C. Y. (2013) *Natural products: an update*. Nat. Prod. Rep. 30, 760–781.

Clardy, J., Bauld, S. M., and Cragg, G. M. (2012) Antimicrobial drug discovery. Nat. Rev. Drug Discov. 11, 827–842.

Fischbach, M. A. and Walsh, T. M. (2006) The role of fungi and bacteria in the production of new classes of antibiotics. Annu. Rev. Microbiol. 60, 363–385.

Gunatilaka, A. A. (2006) Polyketides: an overview of biology, chemistry, and potential industrial applications. In *The Polyketide Compendium* (eds. A. O. P. Patel, N. Patel and A. Hughes), pp. 3–78. Wiley-Blackwell, Oxford.

Klayman, J. L., Glauser, P. D., and Reincke, J. (1984) *Endophytic fungi*—sources of new therapeutic agents. In *New Antimicrobial Agents and Chemotherapies*, pp. 139–185. American Society for Microbiology, Washington, DC.

Newman, D. J. and Cragg, G. M. (2012) Natural products: still a significant source of new drugs or drug leads. *Bioorg. Med. Chem. Lett.* 22, 154–159.

Qin, Y. and Xu, C. (1998) Traditional Chinese medicine and contemporary Chinese medicine (TCM and CCM). *J. Ethnopharmacol.* 64, 93–103.

Wang, X., Lin, C., Zhao, X., Zhou, X., Lu, J., and Wang, X. (2012) Recent advances in the discovery of bioactive natural products from Chinese medicines. In *Recent Advances in Natural Products and Drug Discovery* (eds. F. Chen and T. Chen), pp. 1–37. Bentham Science Publishers, Sharjah, U.A.E.
chemistry, genome screening or a combination of these methods (Alvin et al., 2016).

We previously reported the bioactivity and biosynthetic potential of culturable endophytes from TCM plants used for the treatment of cancer (Miller et al., 2012a,b). As part of that screening process, we found that the organic extract of F53 from the TCM plant T. yunnanensis collected from Yunnan, China, displayed the highest cytotoxicity of twenty unique plant endophytes screened against human multiple myeloma RPMI-8226 cells (Miller et al., 2012b). Screening of genomic DNA using robust PCR methods revealed the presence of polyketide synthase (PKS) genes (Miller et al., 2012b) and this, coupled with the bioactivity displayed by the extract, led us to investigate this organism further.

In the present study, we sequenced and mined the genome of F53 in order to comprehensively assess its potential for the production of specialized (secondary) metabolites. This analysis revealed the presence of 35 putative biosynthesis gene clusters (BGCs), including a unique azaphilone BGC that subsequently guided the discovery of the novel specialized metabolite 1 (Fig. 1). The in silico characterization of the liquipulone (lij) BGC as well as the isolation, structural elucidation and bioactivity of the corresponding novel azaphilone are described below.

Results

Phylogenetic analysis of F53

Preliminary BLASTn analysis of the ribosomal internal transcribed spacer (ITS) and large subunit (LSU) sequences of F53 revealed similarity to species within the order Muyocopronales (Dothideomycetes). The consensus tree of the multi-locus analysis placed F53 within a clade representative of the genus Muyocopron (Muyocopronales, Dothideomycetes) with its closest relative being Muyocopron atromaculans (MUC 34983) (Fig. 2).

F53 genome assembly and specialized metabolite biosynthesis gene cluster prediction

The genome assembly produced by SOAPdenovo, using a k-mer value of 35, resulted in a 38.52 Mbp assembly consisting of 778 scaffolds and 2554 unscaffolded contigs with a N50 of 341 577 and a maximum contig length of 42 378 bp. The SOAPdenovo assembly was chosen for further analyses because it produced more scaffolds, which were also longer than those produced by Velvet. The average GC content of the assembled genome was 40% with 30X coverage. The draft genome for F53 was deposited into GenBank and is available online (accession no. KT874412).

The fungiSMASH 5.1.0 (Blin et al., 2019) annotation revealed 35 putative specialized metabolite BGCs within the F53 genome, including type I iterative PKS (iPKS) BGCs, including a tandem PKS BGC, incorporating two type I iPKSs; a highly reducing PKS (HR-PKS) and a non-reducing PKS (NR-PKS) BGC (scaffold375); non-ribosomal peptide synthase (NRPS) BGCs, PKS-NRPS hybrid BGCs, terpene BGCs, betalactone BGCs and an indole BGC (Fig. S2).

Azaphilone biosynthetic gene cluster (lij) discovery

Several fungal polyketide synthase (PKS) products with azaphilone-type structures have been reported in the literature, including azanigerone A 4 (Zabala et al., 2012) and chaetoviridin A 3 (Winter et al., 2012; Winter et al., 2015; Sato et al., 2016) (Fig. 1), which were shown to be assembled via a convergent pathway in which a dissociated acyl transferase functions to release a diketide from its PKS and transfer it to a PKS-derived acyl acceptor (Xie et al., 2009; Meehan et al., 2011). A phylogenetic comparison of the ketoacyl synthase (KS) domains of known azaphilone PKSs with those identified in the F53 genome revealed a ~35 kb cluster, comprised of eight open reading frames, with high homology to the asperfuranone (afo), azanigerone (aza) and chaetoviridin (caz) BGCs (Fig. 3). Comparison of the domain architecture of the aza (Zabala et al., 2012) and caz (chaetoviridin) (Winter et al., 2012; Winter et al., 2015; Sato et al., 2016) PKSs with that of the PKSs from the candidate F53 BGC showed that it was potentially capable of producing an azaphilone. Significantly, this cluster was also the only tandem iPKS BGC identified from the endophyte and has thus been designated lij (Fig. 4; Table 1).

Careful analysis of the proposed lij BGC revealed a number of genes coding for enzymes whose function can be assigned based on generic azaphilone biosynthesis. Along with the structure of the isolated compound 1 and its proposed biosynthesis pathway, compared to known pathways, this information has allowed a manual curation of the proposed lij BGC (Fig. 4). The lijA gene, which is located on the end of the assembled (~400 kb) scaffold, appeared to be truncated as it lacked an essential phosphopantetheine (PP) binding site. The missing acyl carrier protein (ACP) was identified on an overlapping contiguous fragment via comparison with the homologous lovastatin reducing PKS gene from
Acremonium chrysogenum ATCC 11550, a producer of the β-lactam antibiotic cephalosporin C (Terfehr et al., 2014). Confirmation of the overlap between fragments of the two scaffolds was provided by PCR of the genomic locus.

Isolation and structure elucidation of the novel azaphilone 1

Ethyl acetate extracts of F53 were fractionated using column chromatography (Fig. S4) and screened for new cytotoxic azaphilone, lijiquinone 1.

Fig. 1. The novel azaphilone 1 and structurally related polyketides 2-7.
Cytotoxicity against RPMI-8226 cells. Compound 1 (6 mg l⁻¹) was isolated from a bioactive fraction corresponding to a single peak from C₁₈ reversed phase semi-preparative HPLC, eluting at 15.2 min. It was examined using ¹H NMR and was determined to be sufficiently pure for structural elucidation.

Fig. 2. Combined ITS and LSU phylogram of F53 within the order Muyocopronales. F53 is bolded and coloured red. Pleosporales spp. were used as the outgroup. Evolutionary relationships were determined by Bayesian inference (BI) analysis using a GTR + G substitution model. Branch length indicates inferred divergence of nucleotide sequences. Node labels indicate BI posterior probabilities (percentage) where values >50% were considered significant.

Fig. 3. Phylogeny of F53 KS domains in relation to those that produce azaphilones. KS domains from F53 derived PKSs are in bold with LjE coloured red. Evolutionary relationships were determined by Bayesian inference (BI) analysis using a LG + I + G substitution model. Branch length indicates inferred divergence of amino acid sequences. Node labels indicate BI posterior probability (percentage) where values >50% were considered significant.
The polyketide 1 was isolated as a yellow solid. High-resolution fourier transform mass spectrometry (HRFTMS) displayed molecular ions at m/z 385.16470 [M + H]+ (Δ 0.3528 ppm) and m/z 383.15086 [M-H]- (Δ 2.1924 ppm), in positive and negative ionization modes respectively, indicating a molecular formula of C22H24O6, which required 11 degrees of unsaturation.

The bulk structure of 1 was elucidated using standard NMR methods (Figs S5–S10). The 13C NMR spectrum of 1 showed 22 carbon unique resonances, 14 of which were shown in the HSQC spectrum to be protonated (Table 2). The 1H NMR spectrum of 1 revealed 24 protons which were assigned in concert with the HSQC data, to four methyls, two methylenes and eight methines. Careful analysis of both the COSY and HMBC spectra allowed the assignment of three partial fragments which were ultimately assembled to yield the azaphilone polyketide 1. Key HMBCs observed from H-4 to C-10 and from H-10 to C-4 established the attachment of the methylcyclohexenyl ring fragment to the isochromene core at C-3, and thus, the methyl butanoate fragment was joined to the quaternary C-7 resonating at ~83.9 ppm. While HMBCs could not confirm this attachment, the 13C NMR shift for C-7 is consistent with corresponding carbons in the azanigerones series which occur between 83.1 and 84.2 ppm (Zabala et al., 2012).

Supporting our NMR and mass spectrometry-based structure assignment are recent examples of azaphilones in the literature including azanigerone A from the engineered strain Aspergillus niger T1 (Zabala et al., 2012), multiformin A (Quang et al., 2005b) and cohaerin B (Quang et al., 2005a) (Fig. 1). 4, 5 and 6 represent a distinct class of azaphilone that possess a cyclohexenone side group attached to the core bicyclic.

### Table 1. Putative functions of lij cluster biosynthetic gene products.

| Protein | Amino Acids | Putative function | BLASTP match | Accession No. | Ident. (%) | Cover. (%) |
|---------|-------------|------------------|--------------|---------------|------------|------------|
| LijA    | 2223        | Reducing PKS (KS-AT-DH-cMT-ER-KR-ACP) | Type I iterative polyketide synthase, Pseudogymnoascus sp. WSF 3629 | OBT39594.1 | 73 | 100 |
| LijB    | 284         | Hydrolase        | Citrinin biosynthesis oxidoreductase CtnB, Periconia macrospinosa | PVH94008.1 | 64 | 94 |
| LijC    | 282         | Dehydrogenase    | Putative short-chain dehydrogenase/reductase, M. acridum COMa | XP_ | 84 | 98 |
|         | 007816049.1 |                  |              |               |            |            |
| LijD    | 457         | Hydroxylase      | Putative salicylate hydroxylase, M. acridum COMa | XP_ | 84 | 98 |
|         | 007816050.1 |                  |              |               |            |            |
| LijE    | 2685        | Non-reducing PKS (SAT-KS-AT-PT-ACP-cMT-R) | Type I iterative polyketide synthase, Pseudogymnoascus sp. WSF 3629 | OBT39041.1 | 81 | 99 |
| LijF    | 600         | Oxidase          | Oxidase cueO precursor, Aureobasidium melanogenum CBS110374 | KEQ59244.1 | 59 | 92 |
| LijG    | 483         | Aldehyde dehydrogenase | Aldehyde dehydrogenase, Glonium stellatum | XP_ | 53 | 98 |
|         | 007583778.1 |                  |              |               |            |            |
| LijH    | 321         | Acyl transferase | Carbohydrate esterase, Glonium stellatum | OCL03758.1 | 58 | 98 |

The polyketide 1 was isolated as a yellow solid. High-resolution fourier transform mass spectrometry (HRFTMS) displayed molecular ions at m/z 385.16470 [M + H]+ (Δ 0.3528 ppm) and m/z 383.15086 [M-H]- (Δ 2.1924 ppm), in positive and negative ionization modes respectively, indicating a molecular formula of C22H24O6, which required 11 degrees of unsaturation.

The bulk structure of 1 was elucidated using standard NMR methods (Figs S5–S10). The 13C NMR spectrum of 1 showed 22 carbon unique resonances, 14 of which were shown in the HSQC spectrum to be protonated (Table 2). The 1H NMR spectrum of 1 revealed 24 protons which were assigned in concert with the HSQC data, to four methyls, two methylenes and eight methines. Careful analysis of both the COSY and HMBC spectra allowed the assignment of three partial fragments which were ultimately assembled to yield the azaphilone polyketide 1. Key HMBCs observed from H-4 to C-10 and from H-10 to C-4 established the attachment of the methylcyclohexenyl ring fragment to the isochromene core at C-3, and thus, the methyl butanoate fragment was joined to the quaternary C-7 resonating at ~83.9 ppm. While HMBCs could not confirm this attachment, the 13C NMR shift for C-7 is consistent with corresponding carbons in the azanigerones series which occur between 83.1 and 84.2 ppm (Zabala et al., 2012).

Supporting our NMR and mass spectrometry-based structure assignment are recent examples of azaphilones in the literature including azanigerone A from the engineered strain Aspergillus niger T1 (Zabala et al., 2012), multiformin A (Quang et al., 2005b) and cohaerin B (Quang et al., 2005a) (Fig. 1). 4, 5 and 6 represent a distinct class of azaphilone that possess a cyclohexenone side group attached to the core bicyclic.
A ROESY experiment was used to determine the trans relative stereochemistry of the methyl group (C16) and the pyranquinone around the cyclohexenone ring, which was assigned based on the lack of a ROESY between H-10 and H-11 despite them showing strong COSYs. Key correlations indicating the close proximity between H-10 (δ 3.09) and H-2′ (δ 2.54), H-2′ and the H-3 methyl protons (δ 1.08), and H-2′ and H-5′ (δ 1.20) suggest that the cyclohexenone ring and the methyl butanoate are folded under the relatively planar azaphilone core. The absolute stereochemistries of C-9 and C-2′ together with that of C-7 as depicted are based on biosynthetic precedents (29, 32, 34).

Bioactivity of 1

The novel compound 1 was evaluated for biological activity in vitro and displayed moderate cytotoxic and antifungal activity. A CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS) (Sekhon et al., 2008) showed that 1 was cytotoxic against human multiple myeloma RPMI-8226 cells (IC_{50} = 129 µM) but not CHO-K1 cells (IC_{50} > 300 µM) at the concentrations tested. Compound 1 also showed antifungal activity against C. albicans (IC_{50} = 79 µM) and C. albidos (IC_{50} = 141 µM) but no significant activity (IC_{50} > 300 µM) against the Gram-positive bacterium Bacillus subtilis. In light of these data, 1 was submitted to the US National Cancer Institute for screening against the 60 cancer cell line panel (Shoemaker, 2006). At concentrations of 10 µM, 1 displayed highest activity against renal cancer cells (UO-31) and non-small lung cancer cells (NCI-H522), inhibiting growth by approximately 35 and 25 per cent respectively. Further anticancer testing has not been conducted on 1.

Discussion

This work demonstrates that endophytes of traditional medicinal plants, as well as the plants themselves, are valid targets for the discovery of novel natural products. Our central hypothesis that endophytes are potentially responsible for the medicinal properties of some plants is based on the observation that many clinically relevant natural products are either polyketides or non-ribosomal peptides, and compounds belonging to these structure classes are typically of microbial origin. Most plants are home to diverse communities of endophytes, and thus, it is imperative to prioritize these microbes for further investigation. Here, we employed an innovative microbial strain prioritization protocol based on the bioactivity of crude culture extracts as well as the presence of polyketide and non-ribosomal peptide biosynthesis genes (Miller et al., 2012a; Miller et al., 2012b). As a result, we identified the novel fungal endophyte F53. Further analysis of this organism’s genetic architecture and potential for specialized metabolite biosynthesis guided the discovery of the novel azaphilone 1.

Analysis of the multi-locus phylogeny of F53 allowed its placement within the genus Muyocopron where its closest known relative is M. atromaculans (MUCL 34983) (Fig. 2). This placement was possible due to the recent re-evaluation of the Mycoleptodiscus genus and its subsequent placement within the order Muyocopronales (Dothideomycetes) (Crous et al., 2018; Hernández-Restrepo et al., 2019). The placement of F53 in the genus Muyocopron is supported by the fact that most endophytes belong to the classes Sordariomycetes or Dothideomycetes, with this genus belonging to the later class (Rodriguez et al., 2009; Schoch et al., 2009). Although some species within the Muyocopronales order are plant parasites (Hernández-Restrepo et al., 2019), endophytic species have been isolated, including

Table 2. NMR data of 1 recorded in CDCl₃.

| Position | δ_C | δ_H (mult., J/Hz) | COSY | HMBC (δ_C)(δ_H,δ_C) |
|----------|-----|------------------|------|----------------------|
| 1        | 154.3 | 7.91 (s) | – | C-3, C-4a, C-8, C-8a |
| 3        | 158.0 | – | – | – |
| 4        | 113.6 | 6.25 (brs) | – | C-3, C-4a, C-5, C-8a, C-10 |
| 4a       | 143.0 | – | – | – |
| 5        | 107.2 | 5.65 (d, 1.0) | – | C-4, C-7, C-8a |
| 6        | 194.3 | – | – | – |
| 7        | 83.9 | – | – | – |
| 8        | 193.0 | – | – | – |
| 8a       | 115.3 | – | – | – |
| 9        | 22.0 | 1.59 (s) | – | C-6, C-7, C-8 |
| 10       | 59.6 | 3.09 (d, 12.5) | H-11 | C-3, C-4, C-11, C-15, C-16 |
| 11       | 32.8 | 2.60 (m) | H-10, H-12b, H-16 | C-10, C-13 |
| 12       | 34.0 | 2.63 (brt, 5.3) | 2.25 (ddt, 18.8,10.6, 2.6) | H-13, H-14 | C-10, C-11, C-13, C-14 |
| 13       | 150.7 | 7.10 (ddt, 10.0, 6.0, 2.1) | H-12a, H-12b, H-14 | C-11, C-15 |
| 14       | 129.0 | 6.16 (dd, 10.1, 2.4) | H-12a, H-12b, H-13 | C-10, C-12 |
| 15       | 194.7 | – | – | – |
| 16       | 19.8 | 1.08 (d, 6.5) | H-11 | C-10, C-11, C-12 |
| 1’       | 178.4 | – | – | – |
| 2’       | 39.9 | 2.54 (m) | H-3a’, H-5’ | C-1’, C-3’, C-4’, C-5’ |
| 3’       | 26.6 | 1.75 (dt, 13.7, 7.4) | 1.51 (m) | H-2’, H-3b’ H-4’ | C-1’, C-2’, C-4’, C-5’ |
| 4’       | 11.3 | 0.97 (l, 13.7) | H-3a’, H-3b’ | H-4’ | C-2’, C-3’ |
| 5’       | 16.3 | 1.20 (d, 7.4) | H-5’ | C-1’, C-2’, C-3’ |
Mycoplastodiscus endophytica (Tibpromma et al., 2018) and Muyocoron garethronesii (Tibpromma et al., 2016) from Pandanaceae spp.

A bioinformatics-based analysis of the F53 genome, employing a number of freely available tools to assemble and interrogate sequence data, allowed the identification of 35 putative specialized metabolite BGCs. This number could be inaccurate as draft genomes can display gene cluster fragmentation across multiple contigs, which can cause in silico specialized metabolite analysis tools, such as antiSMASH, to detect extra clusters from separated fragments or not detect clusters that are too fragmented (Blin et al., 2013). Our study confirms that endophytes are potentially rich sources of specialized metabolites, compared to other microbes. Previous studies have identified a similar number of specialized metabolite BGCs encoded within the genomes of the Oryza granulata (Chinese wild rice) endophyte Harpophora oryzae (37) (Xu et al., 2014) and the Bracharia brizantha (bread grass) endophyte Sarocladium brachiariae (34) (Yang et al., 2019), while the larger (52 Mb) genome of the Camellia sinensis (tea) endophyte, Pestalotiopsis fici, encodes 74 putative BGCs (Wang et al., 2015). Uniquely encoded within the F53 genome is a tandem iterative type I PKS pathway (lijk), which is seemingly capable of assembling an azaphilone. In order to verify the role of this cluster in the biosynthesis of an azaphilone, the KS domains of all F53 derived PKSs were compared to those of known azaphilone producing PKSs. It was revealed here that the tandem PKSs in question, LijA and LijE, shared a close phylogenetic relationship to PKSs of the asperfurane 2 (Chiang et al., 2009), chaetoviridin A 3 (Winter et al., 2012) and azanigerone A 4 (Zabala et al., 2012) pathways (Figs 1 and 3). This result pushed us to interrogate crude extracts of F53 for the azaphilone produced by this cluster.

Fractionation of a crude extract of F53 using HPLC (Fig. S4) allowed isolation of the novel azaphilone 1, which was structurally characterized via NMR. While both the unusual methylcyclohexenyl and butanoate pendant moieties are known structural features of fungal natural products, 1 represents the only example of an azaphilone-type compound possessing both of these. The presence of a cyclohexenone ring places 1 in a rare class of azaphilone, including multiformin A 5 (Quang et al., 2005b) and cohaerin B 6 (Quang et al., 2005a). Azaphilones are a well-known family of natural products; however, there is still a need to understand their biosynthesis at the genetic level (Gao et al., 2013).

Preliminary identification of the azaphilone gene cluster lij followed by isolation of the novel azaphilone 1 lead us to believe that both the molecule and pathway are connected. We propose that 1 is biosynthesized in a convergent manner analogous to both 3 and 4 assembly (Winter et al., 2012; Zabala et al., 2012; Winter et al., 2015; Sato et al., 2016), by an iPKS in which the products of the highly reducing PKS (HR-PKS) LijA, and a modified product of the non-reducing PKS (NR-PKS) LijE, combine to form 1 (Fig. 5). Central to the assembly of tandem iPKS products is a disassociated acyl transferase responsible for the convergence of the pathway. In 1 biosynthesis, the lij gene product LijH, a member of the fatty acid acyl transferase-like protein subfamily, is likely to perform this important role.

LijE is responsible for the assembly of the azaphilone core, and though the identity of the starter unit is still unknown, it is plausible that LijE is primed by acetate for subsequent chain elongation and C-methylation (Fig. 5). The timing of the formation of the cyclohexenone ring is unclear; however, we anticipate that it occurs prior to the reductive offloading of the PKS product from LijE. A product template (PT) domain embedded within non-reducing PKSs functions to mediate the cyclization of the nascent polyketide prior to reductive offloading to provide the proposed benzaldehyde intermediate 8 (Fig. 5) (Crawford et al., 2008). Hydroxylation of 8, and the subsequent formation of the pyran and ultimately the pyranoid intermediate 9 (Fig. 5) likely proceeds in a fashion analogous to 4 biosynthesis. This sequence of biosynthesis steps has been confirmed via in vitro studies with the azaH gene product AzaH, which shows similarity to a salicylate monooxygenase (Zabala et al., 2012). Dehydration of 9 provides the acyl acceptor intermediate 10 (Fig. 5) which, upon acyl transferase-mediated acylation with the diketide product of the reducing PKS, yields 1. While the proposed pathway is feasible based on our knowledge of tandem iPKS assembly, this needs to be confirmed by mutagenesis and biochemical experiments. Such experiments could also reveal the timing and formation of the cyclohexenone and the priming of LijE.

Bioassays of 1 revealed cytotoxicity against human multiple myeloma RPMI-8226 cells and antifungal activity against C. albicans and C. albidus. These bioactive properties were previously observed in crude extracts of F53 (Miller et al., 2012b) and can now be attributed to the novel azaphilone 1. The cytotoxic effects of 1 are promising, but it is not as effective as the anticancer drugs paclitaxel (IC50 = 2.5 nM, human breast adenocarcinoma MCF-7 cells) (Liebmann et al., 1993) or camptothecin (IC50 = 51 nM, human ovarian adenocarcinoma SKOV3 cells) (Zhao et al., 1997). Additionally, the slow growth rate of F53 is a present hinderance to the production of commercial quantities of this compound. Potency and yields could potentially be improved via the heterologous expression and modification of the lij cluster in a suitable host (e.g. Saccharomyces cerevisiae or Aspergillus spp.) as has been achieved for the
production of other natural products, such as alternariol (Chooi et al., 2015) and 2 (Chiang et al., 2013).

Cytotoxicity is common among azaphilones (Gao et al., 2013). One of these compounds, chaetomugilin A 7 (Fig. 1), is biosynthesized by enzymes encoded within a homologous BGC to lij and was found to display moderate cytotoxicity against human P388 (IC_{50} = 8.7 \mu M) and HL-60 (IC_{50} = 7.3 \mu M) cell lines, and selective cytotoxicity against a disease orientated panel of 39 human cancer cell lines (Yasuhide et al., 2008; Gao et al., 2013). The cytotoxicity of 1, as observed in this study, suggests that the anticancer properties of the TCM plant T. yunnanensis could be partially mediated by this compound, in addition to other endophyte- or plant-derived natural products, such as paclitaxel (Li et al., 2001).

Analysis of the F53 genome indicated that this endophytic fungus is capable of synthesizing numerous specialized metabolites. In addition to the proposed lij cluster, 34 other putative BGCs were identified within the genome of this organism. Characterization of these BGCs could guide the discovery of additional novel compounds or reveal the genetic basis for previously isolated compounds, such as the toxin alternariol (Dasari et al., 2013).

In summary, this study has revealed that F53, a fungal endophyte of T. yunnanensis has the genetic potential to produce numerous specialized metabolites, including the novel azaphilone 1. The observed bioactive properties of this endophyte-derived compound may partially explain the documented medicinal properties of the host plant. Our study also highlights the utility of traditional knowledge and genome mining as beacons for drug discovery.

**Experimental procedures**

**General experimental**

Plant collection, endophyte isolation, DNA extraction, PCR protocols and bioassay procedures have been described in detail previously (Miller et al., 2012a; Miller et al., 2012b).

**Phylogenetic analysis of F53**

The rDNA internal transcribed spacer (ITS) region and the 5′-end of the 28S rDNA large subunit (LSU) were amplified from F53 genomic DNA using the ITS5/ITS4 (White et al., 1990) and LROR/LR5 (Vilgalys and Hester, 1990; Bunyard et al., 1994) primer pairs respectively. Ethanol precipitated PCR products were sequenced using a Big Dye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Melbourne, VIC, Australia) according to the manufacturer’s protocol and analysed using an ABI 3730 DNA Analyzer (Applied Biosystems). Analysis was performed at the Ramaciotti Centre for Gene Analysis (UNSW).

A multi-locus phylogenetic analysis was undertaken based on concatenated alignments of ITS and LSU datasets. Alignments were created using published
sequences from 20 taxa prominent within the order Muy-
ocopronales, with 2 species from Pleosporales as out-
group (Fig. S1). These alignments were made using ClustalW (Larkin et al., 2007) in BioEdit (Isis Pharmaceuticals) with default parameters and followed by trimming non-aligned ends manually. The multi-locus dataset was created using Sequence Matrix v1.8 for concatenation (Vaidya et al., 2011) and contained 1,473 characters (658 for ITS and 815 for LSU).

A Bayesian inference tree was obtained using MrBayes v3.2.6 (Ronquist et al., 2012) with the Markov chain Monte Carlo (MCMC) method. Optimal substitution models for phylogenetic inferences were determined using jModelTest v2.1.3 (Posada, 2008) based on the critical Akaike information criterion (AICc) with GTR + G substitution model being the most suitable model. Two parallel chains were run for 500 thousand total generations, with a sample frequency of 250, until the trees converged (standard deviation of split frequencies < 0.01). During burn-in, 25% of the sampled trees were discarded. Tracer v1.6.03 in MrBayes v3.2.6 (Ronquist et al., 2012) was used to confirm runs were long enough to effectively sample the distribution with the above run having effective sample sizes of >200 for all parameters and the potential scale reducing factor (PSRF) of various parameters converging on 1. Trees were visualized using the Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019).

**Genome sequencing**

Total DNA was extracted using a Power Soil DNA isolation kit (Mo Bio). Total DNA yield was determined using the Qubit fluorometer double-stranded DNA BR Kit (Invitrogen, Melbourne, VIC, Australia). Genome sequencing was performed using Genome Analyzer Ixl, with the TrueSeq SBS v4 GA kit. This was performed at the Ramaciotti Centre for Gene Analysis (UNSW). De novo genome assembly was performed with SOAP denovo (Luo et al., 2012). Gene prediction and annotation were performed using the best assembly produced by the software. Ab initio gene prediction was performed using GeneMark-ES (Ter-Hovhannisyan et al., 2008) packaged within the MAKER (Cantarel et al., 2008) genome annotation pipeline. Genbank records were generated for each contig using a custom script. Specialized metabolite BGCs were identified using a combination of 2metDB (Bachmann and Ravel, 2009) and fungiSMASH 5.1.0 (Blin et al., 2019). Both of these software packages use profile Hidden Markov Models (pHMMs) of known biosynthesis gene domains to identify specialized metabolite biosynthesis genes and their domain architecture in query sequences. All specialized metabolite BGCs retrieved were manually checked, and further confirmation of domain architecture was performed using NCBI Conserved Domain Database (CDD) search (Marchler-Bauer et al., 2013). The draft genome for F53 was deposited into GenBank.

**Phylogenetic analysis of the ketoacyl synthase domain of the lij PKS**

A dataset was assembled for phylogenetic analysis consisting of 26 ketoacyl synthase (KS) domain protein sequences inferred from PKS gene homologues. Eighteen sequences were from putative PKSs identified in the F53 genome, 6 sequences were from known azaphilone/tan-
dem iterative PKSs (iPKSs), and 2 sequences were from fungal fatty acid synthase pathways (outgroup sequences). The dataset included morphologically correlated, character-
ized and published sequences. All sequences were initially aligned in a 599 amino acid alignment (Fig. S3) by the multiple sequence alignment program ClustalW (Larkin et al., 2007) with default parameters and followed by trimming non-aligned ends manually.

Bayesian inference trees were obtained using MrBayes 3.2.6 with the Markov Chain Monte Carlo (MCMC) method (Ronquist et al., 2012). A LG + I+G substitution model was used, which was the optimal model as determined by Prottest 3.4.2 (Darriba et al., 2011). Two parallel chains were run 250 thousand total generations with a sample frequency of 250, until the tree converged (standard deviation of split frequencies < 0.01). During burn-in, 25% of the sampled trees were discarded. Tracer v1.6.03 in MrBayes v3.2.6 (Ronquist et al., 2012) was used to confirm runs were long enough to effectively sample the distribution with the above run having effective sample sizes of >200 for all parameters and the potential scale reducing factor (PSRF) of various parameters converging on 1. Trees were visualized using the Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019).

**Large-scale F53 culture conditions and organic extraction**

A 10 ml, 5-day-old starter culture was used to inoculate one litre of malt extract broth (BD Difco), which was incubated for 21 days at 25°C, with shaking at 100 rpm. The fungal cells and broth were extracted with an equal volume (1:1) of EtOAc, and the extract was dried over anhydrous Na2SO4 (Sigma-Aldrich, Sydney, NSW, Australia), filtered and evaporated to dryness in vacuo.

**LC-MS analysis of crude organic extracts containing 1**

Crude organic (EtOAc) extracts of F53 were dissolved in methanol and analysed using an Accela LC system attached to an LTQ OrbitrapXL Mass Spectrometer with a HESI source operating in positive and negative
electrospray modes. Extracts were separated on a reversed phase BEH C18 (50 x 2.1 mm, 1.9 µm) UHPLC column (Waters, Sydney, NSW, Australia) employing an initial solvent gradient of 100% water (0.1% formic acid) for 5 min which was subsequently ramped linearly to 100% acetonitrile over 20 min at a flow rate of 0.4 ml min⁻¹. Data were analysed using Xcalibur software (Thermo Fisher, Melbourne, VIC, Australia).

Fractionation and purification of 1

Crude organic (EtOAc) extracts of F53 (665 mg) were fractionated using a silica solid phase extraction (SPE) cartridge (Grace Davison, Columbia, MD, USA). Fractions were eluted stepwise with DCM, EtOAc, EtOAc: MeOH (1:1), and MeOH and each fraction was tested for cytotoxicity against RPMI-8226 cells. The bioactive EtOAc fraction was separated by HPLC with a Discovery BIO Wide Pore C18, (10 x 250 mm, 5 µm; Supelco, Bellefonte, PA, USA) reversed phase column using a Shimadzu Class VP HPLC system equipped with a Shimadzu SPD-M20A Diode Array Detector. HPLC was performed using a H₂O/MeCN (with 0.5% trifluoroacetic acid) gradient at a flow rate of 3.5 ml min⁻¹. Purification was achieved using a solvent gradient from 50% to 100% MeCN over 20 min after an initial 5 min isocratic elution with 50% MeCN. The cytotoxic fraction containing 1 (6 mg) eluted at 15.2 min under these conditions.

Structural elucidation of 1

Proton (¹H), carbon (¹³C) and 2D (COSY, HMOC, HMBC and ROESY) NMR spectra were recorded on a Bruker Avance III 600 MHz spectrometer (operating at 600 MHz for ¹H and 150 MHz for ¹³C) in CDCl₃ (Cambridge Isotope Laboratories, Tewksbury, MA, USA) and referenced to δ7.27 and 77.0 ppm.

MTS and antimicrobial bioassays

The MTS and antimicrobial bioassays were performed as per Miller et al. (2012b). For each MTS assay, 10⁴ of RPMI-8226 and 5 x 10⁴ of CHO-K1 cells from a log-phase culture were seeded and equilibrated overnight in a sterile flat-bottomed 96-well microtitre plate. Following equilibration, plates were incubated with final extract concentrations between 0.5 and 100 µg ml⁻¹ for 48 h at 37°C, in an atmosphere of 5% CO₂ with 95% humidity. Finally, the CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS) was performed according to the manufacturer’s instructions (Promega, Sydney, NSW, Australia) where the absorbance was measured at 492 nm with a POLARstar microplate reader (BMG Labtech, Mornington, VIC, Australia).

Antimicrobial bioassays were performed by incubating exponential growth phase cultures (OD 0.6 at 600 nm) of B. subtilis ATCC 11774, C. albidos ATCC10666 and C. albicans ATCC 1023 in a sterile flat-bottomed 96-well microtitre plate with final extract concentrations between 0.5 and 200 µg ml⁻¹. Bacterial test plates were incubated for 24 h at 30°C and fungal plates for 48 h at 30°C. At conclusion of the incubation period, the optical density was measured at 600 nm with a SpectraMax 340 plate reader (Molecular Devices, San Jose, CA, USA).

Acknowledgements

This research was funded by the Australian Research Council through a Federation Fellowship awarded to B. A. N. (FF0883440). K. I. M. was supported by the Australia-China Special Fund for S&T Cooperation, and the Australia Endeavour Cheung Kong Programme funded by the Australian Government. We would like to thank staff at the Ramaciotti Centre for Genomics (UNSW) and the Mark Wainwright Analytical Centre (UNSW) for assistance with genome sequencing and the acquisition of spectrometric data respectively.

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

J. W. C., K. I. M., J. A. K., R. C. and B. A. N. designed the research plan. J. W. C., K. I. M. and J. A. K. performed experiments. J. W. C., K. I. M., J. A. K. and R. C. analysed data. All authors contributed to writing the manuscript.

References

Adams, J.D., and Lien, E.J. (2013) The traditional and scientific bases for traditional Chinese medicine: Communication between traditional practitioners, physicians and scientists, In Traditional Chinese Medicine: Scientific Basis for its Use. Adams, J.D., and Lien, E.J. (eds). London, UK: Royal Society of Chemistry, pp. 1–10.

Alvin, A., Kalaizis, J.A., Sasia, B., and Neilan, B.A. (2016) Combined genetic and bioactivity-based prioritization leads to the isolation of an endophyte-derived antmycobacterial compound. J Appl Microbiol 120: 1229–1239.

Aly, A.H., Debbab, A., Kjer, J., and Proksch, P. (2010) Fungal endophytes from higher plants: a proli fi c routine for genome sequencing and the acquisition of spectrometric data respectively.

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

J. W. C., K. I. M., J. A. K., R. C. and B. A. N. designed the research plan. J. W. C., K. I. M. and J. A. K. performed experiments. J. W. C., K. I. M., J. A. K. and R. C. analysed data. All authors contributed to writing the manuscript.

References

Adams, J.D., and Lien, E.J. (2013) The traditional and scientific bases for traditional Chinese medicine: Communication between traditional practitioners, physicians and scientists, In Traditional Chinese Medicine: Scientific Basis for its Use. Adams, J.D., and Lien, E.J. (eds). Lon- don, UK: Royal Society of Chemistry, pp. 1–10.

Alvin, A., Kalaizis, J.A., Sasia, B., and Neilan, B.A. (2016) Combined genetic and bioactivity-based prioritization leads to the isolation of an endophyte-derived antmycobacterial compound. J Appl Microbiol 120: 1229–1239.

Aly, A.H., Debbab, A., Kjer, J., and Proksch, P. (2010) Fungal endophytes from higher plants: a prolific source of phytochemicals and other bioactive natural products. Fungal Divers 41: 1–16.

Aly, A.H., Debbab, A., and Proksch, P. (2013) Fungal endo-

phytes – secret producers of bioactive plant metabolites. Pharmazie 68: 499–505.

© 2020 The Authors. Microbial Biotechnology published by John Wiley & Sons Ltd and Society for Applied Microbiology. Microbial Biotechnology, 13, 1415–1427
Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Concatenated ITS and LSU sequence alignment.
Fig. S2. Specialized metabolite gene clusters detected in the F53 genome.

Fig. S3. KS domain sequence alignment.
Fig. S4. HPLC-UV chromatogram (330 nm) of ethyl acetate extracts from F53.
Fig. S5. HPLC-UV chromatogram (330 nm) of ethyl acetate extracts from F53.
Fig. S6. $^{13}$C NMR spectrum of lijiquinone recorded in CDCl$_3$ at 150 MHz.
Fig. S7. COSY spectrum of lijiquinone recorded in CDCl$_3$.
Fig. S8. HMQC spectrum of lijiquinone recorded in CDCl$_3$.
Fig. S9. HMBC spectrum of lijiquinone recorded in CDCl$_3$.
Fig. S10. ROESY spectrum of lijiquinone recorded in CDCl$_3$.