**Comprehensive chemotaxonomic and genomic profiling of a biosynthetically
talented Australian fungus, *Aspergillus burnettii* sp. nov.**

Cameron L. M. Gilchrist¹, Heather J. Lacey², Daniel Vuong², John I. Pitt², Lene Lange^{3,4},
Ernest Lacey^{2,5}, Bo Pilgaard³, Yit-Heng Chooi^{1*} and Andrew M. Piggott^{5*}

¹*School of Chemistry and Biochemistry, University of Western Australia, Crawley, WA 6009, Australia*

²*Microbial Screening Technologies, Smithfield, NSW 2164, Australia*

³*Center for Bioprocess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, Denmark*

⁴*BioEconomy, Research & Advisory, Karensgeade 5, 2500 Valby, Copenhagen, Denmark*

⁵*Department of Molecular Sciences, Macquarie University, NSW 2109, Australia.*

* Corresponding authors YHC: Ph: +61 8 6488 3041; Email: yitheng.chooi@uwa.edu.au
AMP: Ph: +61 2 9850 8251; Email: andrew.piggott@mq.edu.au

23 **Abstract**

24 *Aspergillus burnettii* is a new species belonging to the *A. alliaceus* clade in *Aspergillus* subgenus
25 *Circumdati* section *Flavi* isolated from peanut-growing properties in southern Queensland, Australia.
26 *A. burnettii* is a fast-growing, floccose fungus with distinctive brown conidia and is a talented
27 producer of biomass-degrading enzymes and secondary metabolites. Chemical profiling of *A.*
28 *burnettii* revealed the metabolites ochratoxin A, kotanins, isokotanins, asperlicin E, anominine and
29 paspalinine, which are common to subgenus *Circumdati*, together with burnettiene A, burnettramic
30 acids, burnettides, and high levels of 14 α -hydroxypaspalinine and hirsutide. The genome of *A.*
31 *burnettii* was sequenced and an annotated draft genome is presented. *A. burnettii* is rich in secondary
32 metabolite biosynthetic gene clusters, containing 51 polyketide synthases, 28 non-ribosomal peptide
33 synthetases and 19 genes related to terpene biosynthesis. Functional annotation of digestive enzymes
34 of *A. burnettii* and *A. alliaceus* revealed overlapping carbon utilisation profiles, consistent with a
35 close phylogenetic relationship.

37 **Keywords**

38 *Aspergillus*
39 Secondary metabolites
40 Chemotaxonomy
41 Carbohydrate active enzymes
42 Biosynthetic gene cluster analysis
43 Molecular phylogeny
44

1. Introduction

Deciphering the nexus between genomic diversity and biochemical diversity is one of the frontiers in microbial biodiscovery. For many decades, bioactivity-guided discovery has dominated microbial natural products research, largely driven by a desire to identify new and more potent “actives” against pharmaceutically and agriculturally relevant targets (Davies, 2006). However, with the advancement of genome sequencing technologies and the development of new and more powerful analytical tools, there is now a far greater appreciation of the vast biosynthetic potential of microorganisms, especially fungi. Biosynthetic gene clusters (BGCs) occupy a significant percentage of fungal genomes and vary enormously, even between closely related species. A typical fungus may contain 50-100 BGCs, yet few organisms produce this number of active compounds in laboratory culture conditions. Indeed, a facile analysis of actives as a function of total metabolites suggests as many as 95% of metabolites currently have no known function and therefore *ipso facto* represent a window into unexplored biology (Li et al., 2016; Walsh and Tang, 2017). Unravelling the true biological and ecological roles of these natural products is challenging and often requires a multifaceted approach drawing on intimate knowledge of the chemistry, biology and ecology of the producing organism (Piggott and Karuso, 2004; Chooi and Solomon, 2014; Piggott and Karuso, 2016).

Among the filamentous fungi, the genus *Aspergillus* is particularly endowed with biosynthetic capabilities to generate structurally diverse natural products including clinical drugs, e.g. the cholesterol-lowering polyketide-derived lovastatin (Endo, 2004) and antifungal lipopeptide echinocandin (Denning, 2003), and pharmaceutical leads, e.g. the antiangiogenic meroterpenoid fumagillin (Lefkove et al., 2007) and anticancer alkaloid phenylahistin (Kano et al., 1997). The biosynthetic talent of the Aspergilli is further demonstrated by recent genomic analyses, which unveiled the large number of BGCs encoded in their genomes (de Vries et al., 2017; Vesth et al., 2018). Therefore, systematic chemotaxonomic and genomic studies of rare or novel *Aspergillus*

70 species is a fruitful strategy for discovering novel bioactive natural products, as exemplified by our
71 recent work (Lacey et al., 2016; Pitt et al., 2017; Chaudhary et al., 2018; Li et al., 2019).

72
73 In addition to bioactive secondary metabolites, Aspergilli are also important producers of industrial
74 enzymes with applications in food, textile, recycling and other industries (Cairns et al., 2019), many
75 of which belong to various carbohydrate-active enzyme (CAZyme) families. Fungal genomes possess
76 an array of CAZyme genes (typically 400-1000 CAZyme genes, as per the JGI MycoCosm database).
77 Having an armament of enzymes capable of breaking down different carbon sources is critical for
78 fungal growth on diverse substrates and is essential for species fitness and competitiveness (Barrett
79 et al., 2020). Indeed, possessing such a flexible set of enzymes is one of the main reasons why ruderal
80 Aspergilli dominate in unpredictable and disturbed habitats. An understanding of CAZymes is also
81 relevant for elucidating the evolution of the microbial interaction secretome, composed of secreted
82 metabolites and enzymes. Several Aspergilli, such as *Aspergillus niger*, are important for the
83 production of organic acids and serve as hosts for recombinant production of secreted enzymes for
84 use in industrial biotechnology (Lange, 2017).

85
86 Despite their biotechnological potential, Aspergilli are known to produce several important
87 mycotoxins, such as aflatoxins and ochratoxins that are harmful to human health and threaten food
88 security. Contamination of post-harvest food crops with mycotoxins contributes to significant losses
89 in crop productivity and consequently economy in many countries (Pitt and Hocking, 2009). One of
90 the major affected crops is peanut, which is often contaminated with two aflatoxigenic *Aspergillus*
91 species, *A. flavus* and *A. parasiticus*. One approach for reducing aflatoxin occurrence in peanuts is
92 via competitive inhibition or exclusion by nontoxigenic *Aspergillus* strains (Pitt and Hocking, 2006).
93 Therefore, the interactions between *A. flavus* and *A. parasiticus* and native Aspergilli already present
94 in the soil of the cultivation area are of interest in the search for such biocontrol strains. Herein, we

95 describe a novel species *Aspergillus burnettii*, which was isolated during an ecological survey of
96 *Aspergillus* species in the South Burnett peanut-growing region of Southeast Queensland in 1982. *A.*
97 *burnettii* was recovered from arable soil at Coalstoun Lakes that had previously been under peanut
98 (*Arachis hypogaea*) cultivation.

99

100 We recently reported a novel family of bolaamphiphilic antifungal compounds, the burnettramic
101 acids, from *A. burnettii*, along with the biosynthetic gene cluster encoding their biosynthesis, (Li et
102 al., 2019). However, the broader morphotaxonomic, chemical and genomic profiles of the fungus
103 have not been thoroughly investigated. In the present paper, we present a comprehensive analysis of
104 the phenotype, genotype and chemotype of *A. burnettii* to fully characterise this new species.
105 Additionally, we updated the draft genome assembly and functional annotation of *A. hancockii* FRR
106 3425, which we recently described (Pitt et al., 2017), for comparison with *A. burnettii* along with *A.*
107 *flavus*. Chemical analyses revealed a number of structurally diverse metabolites produced by *A.*
108 *burnettii*, including several unique metabolites and metabolites shared with closely related *Aspergilli*.
109 However, this number pales in comparison to the biosynthetic potential encoded in the genome. We
110 also investigated the interplay between indigenous *A. burnettii* and three strains of *A. flavus* isolated
111 from the same environment. This provided an opportunity to observe the adaptive strategies of native
112 microbial species in agricultural systems impacted by plant monoculture and their pathogens.

113

114

115 **2. Materials and Methods**

116 **2.1 Morphotaxonomical characterisation**

117 Isolates of *Aspergillus burnettii* were grown for 7 days on the standard regimen (Pitt and Hocking,
118 2009) of Czapek yeast extract agar (CYA), malt extract agar (MEA) and glycerol nitrate agar (G25N)
119 at 25 °C, and CYA at 37 °C, and morphological descriptions were prepared. Colours mentioned are
120 taken from the Methuen Handbook of Colour (Kornerup and Wanscher, 1978). Colonies on CYA and
121 MEA were photographed after growth for 7 days on CYA and MEA, and photomicrographs were
122 prepared from colonies on CYA after 7 days incubation at 25 °C. Cultures were accessioned into the
123 culture collection at Microbial Screening Technologies, Smithfield, NSW, as MST FP2249, and the
124 collection at CSIRO Agriculture and Food, North Ryde, NSW, as FRR 5400. The type was deposited
125 with NSW Department of Agriculture Herbarium, Orange, NSW, under accession code DAR 84902.

127 **2.2 Carbon utilisation profiling**

128 *A. burnettii* MST-FP2249, grown on potato dextrose agar (Pitt and Hocking, 2009) was used to
129 inoculate 5 mL of liquid wheat bran medium (5% wheat bran (Finax, Denmark) in deionised water).
130 The culture was incubated at 20 °C, 175 rpm on an orbital shaker for 7 days. The supernatant was
131 harvested by centrifugation. The supernatant was tested on 11 AZurine Cross-Linked (AZCL)
132 substrates (Megazyme, Ireland) for establishing an enzyme activity profile of *A. burnettii* grown on
133 liquid wheat bran medium. The AZCL substrates were prepared as earlier described (Pitt et al., 2017).
134 Enzyme activity was confirmed by a blue halo around the sample well, indicating the presence of
135 active enzymes, which can break down the specific AZCL substrate, thereby releasing the blue
136 soluble dye. Measurable blue halos ranged from 0.5 mm to 12 mm radius, providing a sensitive, semi-
137 quantitative assay.

138

139 **2.3 Secondary metabolite profiling and characterisation**

140 *2.3.1 General Experimental Details*

141 NMR spectra were obtained on a Bruker Avance DRX600 spectrometer in the solvents indicated and
142 referenced to residual signals in the deuterated solvents. High resolution electrospray ionisation mass
143 spectra (HRESIMS) were obtained on an Agilent G6538A Q-TOF mass spectrometer by direct
144 infusion. Electrospray ionisation mass spectra (ESIMS) were acquired on an Agilent 1260 UHPLC
145 coupled to an Agilent G6130B single quadrupole mass detector. Chiroptical measurements ($[\alpha]_D$)
146 were obtained on a JASCO P-1000 polarimeter in a 100 × 10 mm cell. UV-vis spectra were acquired
147 in MeCN on a Varian Cary 300 spectrophotometer in a 10 × 10 mm quartz cell. Analytical HPLC
148 was performed on a gradient Shimadzu HPLC system comprising a LC-10AT VP gradient
149 chromatograph, SPD-M10A VP diode array detector and SCL-10A VP system controller. The
150 column was an Alltima C₁₈ “rocket” format column (100 Å, 53 × 7 mm, 3 µm; Grace Discovery,
151 Deerfield, IL, USA) eluted with a 3 mL/min gradient of 10–100% MeCN/H₂O (+ 0.01% TFA) over
152 7 min. Preparative HPLC was performed on a gradient Shimadzu HPLC system comprising two LC-
153 8A preparative liquid pumps with a static mixer, SPD-M10AVP diode array detector and SCL-
154 10AVP system controller with a standard Rheodyne injection port. The column was a preparative
155 Vydac C₁₈ (150 × 50 mm, 5 µm; Grace Discovery) eluted isocratically at 60 mL/min. Further
156 purification was undertaken using a preparative Platinum EPS C₁₈ column (150 × 22 mm, 5 µm;
157 Grace Discovery) eluted isocratically at 10 mL/min.

158

159 *2.3.2 Growth Optimisation*

160 *A. burnettii* was cultivated on a range of agar, liquid and grain-based media to identify optimal
161 conditions for production of secondary metabolites. Czapek-dox agar (CZA), malt extract agar
162 (MEA), yeast extract sucrose agar (YSA), oatmeal agar (OMA), and casein glycerol Agar (CGA),
163 were prepared according to the recipes in Table S25. Liquid cultures were based on the agar recipes,

164 omitting agar. The grains, pearl barley (PB) and cracked wheat (BL), and the rices, jasmine (JR) and
165 basmati (BS), were prepared by hydration (50 g with 30 mL water in a 250 mL flask) during
166 sterilisation (120 °C for 40 min). The agar and grains were inoculated with a suspension of fungal
167 spores and incubated at 24 °C for 14 days. The cultures were sub-sampled (1 g) and extracted with
168 methanol (2 mL) 1 h on a wrist shaker, centrifuged ($15,700 \times g$ for 3 min) and analysed by HPLC
169 (Figure S1). The HPLC traces were accessioned into our in-house co-metabolite databases, COMET
170 (Lacey and Tennant, 2003) and NovaC. The metabolite profile of each extract was analysed for
171 retention time, UV-vis and MS fit against known standards isolated in this project (Table S26).

172

173 2.3.3 Preparative Cultivation and Isolation

174 A spore suspension of *A. burnettii* from a 7-day-old Petri plate was used to inoculate 40×250 mL
175 Erlenmeyer flasks, each containing 80 g of the hydrated and autoclaved pearl barley. The barley was
176 incubated at 24 °C for 21 days, by which time the culture had grown extensively throughout the grain
177 reaching maximal metabolite productivity. The grains from individual flasks were pooled into 2×5
178 L Erlenmeyer flasks for processing. The pooled grains (3.2 kg) were extracted with acetone (2×3
179 L), evaporated to an aqueous residue *in vacuo* (132 g). The residue was partitioned against ethyl
180 acetate (2×2 L) to give crude organic (15 g) and aqueous (116 g) extracts. The crude organic extract
181 was redissolved in 90% aqueous methanol (300 mL) and partitioned over hexane (2×300 mL) to
182 remove the lipids and provide an enriched fraction (9.4 g). Sephadex LH-20 chromatography ($24 \times$
183 5.5 cm column; MeOH; 20-mL fractions) and analytical HPLC analysis resolved the enriched
184 material into three pooled fractions (Fractions 12-20, 21-25 and 26-30).

185

186 Fraction 12-20 (5.3 g) was fractionated by preparative reversed phase HPLC (Vydac C₁₈, isocratic
187 60/40 MeCN/H₂O, 60 mL/min) to yield ochratoxin A (t_R 6.3 min; 401 mg), (+)-kotanin (t_R 8.3 min;
188 10.8 mg) and a mixed fraction (t_R 12.1 min, 321 mg), which was resolved by further preparative

189 reversed phase HPLC (Platinum EPS C₁₈, isocratic 62.5:37.5 MeCN/H₂O + 0.01% TFA, 10 mL/min)
190 to yield hirsutide (*t_R* 11.7 min; 113.3 mg) and burnettiene A (*t_R* 14.1 min; 27.2 mg).
191
192 Fraction 21-25 (0.73 g) was fractionated by preparative reversed phase HPLC (Vydac C₁₈, isocratic
193 60/40 MeCN/H₂O, 60 mL/min) to afford a mixed polar fraction (35.3 mg), (+)-isokotanin B (*t_R* 7.3
194 min; 75.8 mg), (+)-isokotanin A (*t_R* 8.6 min; 14.9 mg) and anominine (*t_R* 16.6 min; 37.7 mg), together
195 with an uncharacterised red pigment (*t_R* 13.5 min; 1.7 mg). Flushing the column with 100% MeCN
196 afforded a mixed fraction of anominine analogues, which were resolved by preparative reversed phase
197 HPLC (Platinum EPS C₁₈, isocratic 90:10 MeCN/H₂O + 0.01% TFA, 10 mL/min) to yield, paspaline
198 (*t_R* 3.0 min; 3.9 mg) and emindole SB (*t_R* 4.5 min; 6.8 mg). The mixed polar fraction was further
199 purified by silica gel column chromatography using a gradient of 0, 1, 2, 4, 8, 16, 32 and 100% MeOH
200 in CHCl₃ to yield asperlicin E (2.1 mg) in the 2% fraction.
201
202 Fraction 26-30 (0.52 g) was separated by size-exclusion chromatography (Sephadex LH-20, MeOH,
203 28 × 3 cm, 20-mL sub-fractions) and further purified by preparative reversed phase HPLC (Platinum
204 EPS C₁₈, isocratic 65:35 MeCN/H₂O + 0.01% TFA, 10 mL/min) to give 14 α -hydroxypaspalinine (*t_R*
205 12.0 min; 193 mg). The crude aqueous extract was fractionated on a C₁₈ SPE cartridge using a 25%
206 stepwise gradient of MeOH in H₂O. The fraction eluting with 75% MeOH/H₂O was further purified
207 by preparative reversed phase HPLC (Vydac C₁₈, isocratic 40/60 MeCN/H₂O + 0.01% TFA, 60
208 mL/min) to yield burnettide A (*t_R* 5.6 min; 2.1 mg) and burnettide B (*t_R* 8.43 min; 9.8 mg).
209 Burnettramic acid A was recovered in an insoluble fraction at the solvent interface of the aqueous
210 ethyl acetate partition and its purification has been reported elsewhere (Li et al., 2019). For a full
211 isolation scheme, see Figure S2. The metabolites were formally characterised by NMR (Table S1-
212 S11 and Figure S3-S24) and UV-vis (Figure S25-S30) spectroscopy.
213

214 2.3.4 Bioassays

215 Purified metabolites were dissolved in DMSO to provide stock solutions (10 mg/mL or 5 mg/mL
216 depending on the amount of material available). An aliquot of each stock solution was transferred to
217 the first lane of Rows B to G in a 96-well microtitre plate and two-fold serially diluted with DMSO
218 across the 12 lanes of the plate to provide a 2,048-fold concentration gradient. Bioassay medium was
219 added to an aliquot of each test solution to provide a 100-fold dilution into the final bioassay, thus
220 yielding a test range of 100 to 0.05 µg/mL in 1% DMSO. Row A contained no test compound (as a
221 reference for no inhibition) and Row H was uninoculated (as a reference for complete inhibition).

222

223 **CyTOX** is an indicative bioassay platform for discovery of antitumour actives. NS-1 (ATCC TIB-
224 18) mouse myeloma and NFF (ATCC PCS-201) human neonatal foreskin fibroblast cells were each
225 inoculated in 96-well microtitre plates (190 µL) at 50,000 cells/mL in DMEM (Dulbecco's Modified
226 Eagle Medium + 10% fetal bovine serum (FBS) + 1% penicillin/streptomycin (10,000 U/mL / 10,000
227 µg/mL, Life Technologies Cat. No. 15140122), together with resazurin (250 µg/mL; 10 µL) and
228 incubated in 37 °C (5% CO₂) incubator. The plates were incubated for 72 h during which time the
229 positive control wells change colour from a blue to pink colour. The absorbance of each well was
230 measured at 605 nm using a Spectromax plate reader (Molecular Devices).

231

232 **ProTOX** is a generic bioassay platform for antibiotic discovery. In the present study *Bacillus subtilis*
233 (ATCC 6633) and *Staphylococcus aureus* (ATCC 25923) were used as indicative species for Gram
234 positive antibacterial activity. A bacterial suspension (50 mL in 250 mL flask) was prepared in
235 nutrient media by cultivation for 24 h at 250 rpm, 28 °C. The suspension was diluted to an absorbance
236 of 0.01 absorbance units per mL, and 10 µL aliquots were added to the wells of a 96-well microtitre
237 plate, which contained the test compounds dispersed in nutrient broth (Amyl) with resazurin (12.5

238 $\mu\text{g/mL}$). The plates were incubated at 28 °C for 24 h during which time the positive control wells
239 change colour from a blue to light pink colour. MIC end points were determined visually. The
240 absorbance was measured using Spectromax plate reader (Molecular Devices) at 605 nm.

241

242 **EuTOX** is a generic bioassay platform for antifungal discovery. In the present study, the yeasts
243 *Candida albicans* (ATCC 10231) and *Saccharomyces cerevisiae* (ATCC 9763) were used as
244 indicative species for antifungal activity. A yeast suspension (50 mL in 250 mL flask) was prepared
245 in 1% malt extract broth by cultivation for 24 h at 250 rpm, 28 °C. The suspension was diluted to an
246 absorbance of 0.005 and 0.03 absorbance units per mL for *C. albicans* and *S. cerevisiae*, respectively.
247 Aliquots (20 μL and 30 μL) of *C. albicans* and *S. cerevisiae*, respectively, were applied to the wells
248 of a 96-well microtitre plate, which contained the test compounds dispersed in malt extract agar
249 containing bromocresol green (50 $\mu\text{g/mL}$). The plates were incubated at 28 °C for 24 h during which
250 time the positive control wells change colour from a blue to yellow colour. MIC end points were
251 determined visually. The absorbance was measured using Spectromax plate reader (Molecular
252 Devices) at 620 nm.

253

254 **2.4 Genome sequencing and analysis**

255 *A. burnettii* was grown in liquid malt extract medium for 7 days with gentle shaking (40 rpm) at room
256 temperature. DNA was extracted from mycelia (200 mg) using the PowerSoil DNA Isolation Kit (Mo
257 Bio Laboratories, Carlsbad, CA). For genomic sequencing, Nextera XT libraries were prepared using
258 Illumina kits (Sample Prep Kit, Illumina, San Diego, CA), then subjected to quality control using a
259 Bioanalyzer 2100 (Agilent Technologies) before whole genome sequencing was carried out using 100
260 bp paired-end Illumina HiSeq 2000.

261 Raw sequencing reads were processed and assembled using the AAFTF pipeline (Stajich and Palmer,
262 2018). Briefly, BBDuk (Bushnell, 2014) was used to trim adapter sequences as well as filter reads for
263 common contaminants and mitochondrial DNA. Reads were then assembled using SPAdes v.3.13.0
264 (Bankevich et al., 2012). Assembly scaffolds containing non-ascomycete sequences identified by
265 sourmash (Pierce et al., 2019), low-coverage scaffolds identified by BWA (Li and Durbin, 2009) and
266 SAMtools (Li et al., 2009), and duplicate scaffolds identified by Minimap2 (Li, 2018), were removed.
267 Genome size was estimated using kmercountexact from the BBTools package (Bushnell, 2014).
268
269 Gene model prediction and functional annotation was performed using the Funannotate v1.5.2
270 pipeline (Palmer and Stajich, 2019). Repetitive elements were identified using RepeatModeler open-
271 1.0.11 (Smit and Hubley, 2015) and masked using RepeatMasker open-4.0 (Smit et al., 2015). Genes
272 were predicted and translated using both AUGUSTUS v3.2.3 (Stanke et al., 2008) trained by BUSCO
273 v2.0 (Simão et al., 2015), and self-trained GeneMark-ES v4.32 (Ter-Hovhannisyan et al., 2008).
274 Functional annotation was performed using InterProScan v5.29-68.0 (Jones et al., 2014), eggNOG-
275 mapper v1.0.3 (Huerta-Cepas et al., 2017), Phobius v1.01 (Käll et al., 2004) and SignalP v4.1
276 (Petersen et al., 2011). Further functional annotations were added via querying the Pfam v31.0 (Finn
277 et al., 2016), BUSCO v2.0 and dbCAN v6.0 (Zhang et al., 2018) databases using HMMER v3.1b2
278 (Mistry et al., 2013), and the MEROPS v12.0 (Rawlings et al., 2014) and UniProt 2018_03 (Bateman
279 et al., 2017) using DIAMOND v0.8.36 (Buchfink et al., 2015). The fully annotated draft genome was
280 deposited on NCBI under accession SPNV000000000 (BioProject: PRJNA516074, BioSample:
281 SAMN10782537).

282

283 **2.5 Molecular phylogenetic analysis**

284 A phylogenetic tree was inferred from combined Internal Transcribed Spacer region (ITS), β -tubulin
285 (BenA), calmodulin (CaM) and second largest RNA polymerase II subunit (RPB2) sequences.

Sequences were obtained from NCBI where available, and otherwise extracted manually via tBLASTn searches against the corresponding genome. Accessions of marker sequences are listed in Table 3. Sequence alignments of each marker were constructed separately using MAFFT v7.273 (Kato and Standley, 2013), trimmed to the first and last non gap-containing columns and then concatenated. Best-fit nucleotide substitution models were chosen using PartitionFinder 2.1.1 (Lanfear et al., 2017), with each marker sequence considered a separate partition. A GTR+I+G model was used for the ITS partition, and a SYM+G model was used for both a merged BenA and CaM partition, as well as the RPB2 partition. Bayesian inference was performed using MrBayes v3.2.7a with 2 independent runs over 1 million generations, each with four Markov chains (3 heated, 1 cold) which were sampled every 1000 generations. Runs were analysed for convergence using Tracer v1.7.1 (Rambaut et al., 2018); effective sample size for all parameters was greater than 200. Posterior probabilities were calculated after discarding the first 25 percent of generations as burn-in. Maximum likelihood trees were constructed using RAxML-NG v0.9.0 (Kozlov et al., 2019) with 1000 bootstrap replicates. Marker-based analysis was performed using a custom Python toolkit (fungphy; <https://www.github.com/gamcil/fungphy>), and trees were visualised using the ETE3 toolkit (Huerta-Cepas et al., 2016).

302

303 **2.6 Biosynthetic gene cluster analysis**

Secondary metabolite biosynthetic gene clusters (BGCs) were predicted using antiSMASH v4.3.0 (Blin et al., 2017) using the arguments ‘--minimal --smcogs --inclusive’. The KnownClusterBlast module was enabled to compare predicted BGCs with those validated and characterised BGCs deposited on the Minimal Information about a Biosynthetic Gene cluster (MIBiG) database (Medema et al., 2015). The conserved functional domains present in PKSs and NRPSs predicted by antiSMASH were further confirmed using synthaser (Gilchrist, 2020a). BGCs predicted by antiSMASH from *A. burnettii* MST FP2249, *A. hancockii* FRR 3425, *A. alliaceus* CBS 536.65, *A. oryzae* 3.042, *A. flavus*

311 AF70 and *A. parasiticus* SU-1 were placed into gene cluster families (GCFs) using the Biosynthetic
312 Genes Similarity Clustering and Prospecting Engine (BiG-SCAPE) (Navarro-Muñoz et al., 2018).
313 Candidate BGCs encoding isolated compounds were identified using cblaster (Gilchrist, 2020b), with
314 pairwise amino acid alignments and visualisation using MUSCLE (Edgar, 2004) v3.8.31 and custom
315 Python scripts (<https://github.com/gamcil/crosslinker>), respectively.

316

317 **2.7 Assignment of CAZyme families**

318 Protein sequences were extracted from the genomes of *A. burnettii* MST FP2249, *A. hancockii* FRR
319 3425, *A. alliaceus* CBS 536.65, *A. oryzae* 3.042, *A. flavus* AF70, *A. parasiticus* SU-1 and analysed
320 for the presence of CAZyme families using the dbCAN 2.0 webserver (Zhang et al., 2018). The total
321 number of copies of specific CAZyme families and subfamilies, as well as a full listing of family
322 assignments per query protein is available in Table S23.

323

324 A more thorough investigation of the potential ability of *A. burnettii* to produce a spectrum of
325 enzymes of specific functions was undertaken using a new sequence analysis method Conserved
326 Unique Peptide Patterns, CUPP (Barrett & Lange, 2019). The CUPP method has been validated,
327 characterised to be on par or beyond state- of-the-art methods for functional annotation with respect
328 to precision, sensitivity and speed. CUPP facilitates assignment of enzymes to protein family and
329 subfamily as well as predicting enzyme function directly from sequence. Based on sharing similar
330 unique peptide patterns, the enzymes can be assigned to CUPP groups, a level of subdivision lower
331 than subfamily, validated to coincide with and hereby facilitate prediction of enzyme function for the
332 CUPP group in question. CUPP was also used to analyse the genome of *A. alliaceus* CBS 536.65 for
333 comparison with *A. burnettii*. An extended table with CUPP results for both species is provided in
334 Table S24.

335

336

337 **3. Results and Discussion**

338 **3.1 Taxonomic Description – *Aspergillus burnettii* Pitt sp. nov**

339 CYA, 25 °C, 7 days: Colonies 55–60 mm diameter, plane, deep and floccose; margins entire;
340 mycelium white; conidial production sparse to moderate, coloured brown, between 4B4 and 5B4;
341 sclerotia characteristically produced on the agar surface at the interface between colonies, initially
342 white, becoming black and rock hard; clear exudate produced, visible at colony interfaces; soluble
343 pigment not produced; reverse uncoloured to pale brown.

344

345 MEA, 25 °C, 7 days: Colonies covering the whole Petri dish, plane, very floccose and deep, but not
346 reaching the Petri dish lid; margins inconspicuous; mycelium white to off-white; conidial production
347 sparse, coloured pale yellow (4A3); exudate and soluble pigment not produced; reverse pale yellow
348 brown.

349

350 G25N, 25 °C, 7 days: Colonies 30–35 mm diameter, lightly floccose, irregularly wrinkled; margins
351 wide, often subsurface, entire; conidial production sparse, pale yellow (4A3); reverse uncoloured.

352

353 CYA, 37 °C, 7 days: Colonies 40–45 mm diameter, plane, velutinous; conidial production heavy,
354 coloured as on CYA at 25 °C or darker brown; exudate and soluble pigment not produced; reverse
355 pale brown.

356

357 Conidiophores borne from aerial hyphae, stipes mostly $750\text{--}1000 \times 9\text{--}11 \mu\text{m}$, with smooth walls;
358 vesicles spherical $50\text{--}70 \mu\text{m}$ diameter, bearing metulae and phialides around the entire circumference;
359 metulae closely packed, $9\text{--}10 \times 2.2\text{--}2.5 \mu\text{m}$; phialides acerose, $9\text{--}10 \times 2.0\text{--}2.2 \mu\text{m}$; conidia spherical
360 or near $2.5\text{--}3.0 \mu\text{m}$ diameter, with thin, smooth walls, borne in radiate heads, in age splitting into 2–

361 6 columns.

362

363 Typification. Holotype DAR 84902 from cultivated soil, Coalstoun Lakes, Queensland, 2004, A.-L.

364 Markovina. Ex-type FRR 5400, MST FP2249, CBS 146237. MycoBank No. MB835453. ITS

365 barcode: MK429758 (alternative markers: *BenA* = MT211761; *CaM* = MT211762; *RPB2* =

366 MT211763).

367

368 Etym.: named for the South Burnett region of Queensland, which includes the small town of

369 Coalstoun Lakes.

370

371 Isolates examined. FRR 5400, MST FP2249 from cultivated soil, Coalstoun Lakes, Queensland,

372 2004, A.-L. Markovina. FRR 6110, MST FP2304, from cultivated soil, Kingaroy, Queensland, 2006,

373 J.I. Pitt. The two isolates were collected from soils 150 km apart, both under cultivation with peanuts

374 (*Arachis hypogea*).

375

Figure 1. *Aspergillus burnettii* MST FP2249 (a) colonies on CYA (left) and MEA (right), 7 days, 25 °C, (b)–(d) fruiting structures, bars 20 µm, (e) conidia, bar 5 µm

380 **3.2 Morphotaxonomy of *Aspergillus burnettii* indicates placement in section *Flavi***

381 Morphologically, the new species *Aspergillus burnettii* produced rapidly growing, loosely textured
382 floccose colonies, lightly sporing on yellow brown fruiting structures indicative of the *A. alliaceus*
383 clade in *Aspergillus* subgenus *Circumdati* section *Flavi* (Figure 1). Species strains within this clade
384 produce bullet shaped sclerotia, a distinctive feature seen elsewhere only in *A. nomius* in section
385 *Flavi*. *A. burnettii* is closely related to *A. alliaceus*, which was isolated from a dead blister beetle in
386 Washington DC, USA (Thom and Church, 1926). *A. alliaceus* differs from *A. burnettii* by the
387 production of yellow to ochre conidia.

388

389 **3.3 *A. burnettii* can utilise a variety of carbon sources**

390 The semi-quantitative AZCL enzyme activity profile indicates *A. burnettii* possesses the ability to
391 utilise a wide spectrum of natural, complex substrates. Eleven substrates were used to test the
392 production of carbohydrate-active enzymes by *A. burnettii* (Table 1). A well-induced culture broth of
393 *A. burnettii*, grown on wheat bran liquid media, was shown to have strongest activity on starch, xylan
394 and arabinoxylan. Good activities were also observed on galactomannan, β glucan and HE-cellulose.
395 Interestingly, only very low activity was observed on xyloglucan, even though this substrate is
396 commonly found in dicot plants. When comparing the enzyme activity profile of *A. burnettii* with the
397 activity profile of *A. hancockii*, significant differences were found. For example, *A. hancockii* showed
398 only limited activity on starch and galactomannan and no activity on galactan (Table 1).

399

400

401 **Table 1.** Enzyme activity profile of liquid culture of *Aspergillus burnettii* MST FP2249.

AZCL Substrate	Enzyme activity	Average zone radius (mm); <i>A. hancockii</i> FRR 3425 activities in parentheses
Amylose	α -Amylase	6 (2)
β -Glucan	β -Glucanase	3 (3)
HE-Cellulose	Endo-1,4- β -glucanase	2 (2)
Galactomannan	Endo-1,4- β -mannanase	4 (1)
Xyloglucan	Xyloglucanase	0.5 (1)
Xylan	Endo-1,4- β -xylanase	6 (5)
Arabinoxylan	Arabinoxylanase	6 (6)
Curdlan and Pachyman	Endo-1,3- β -glucanase	0 (0)
Galactan	Endo-1,4- β -galactanase	2 (0)
De-branched arabinan	Endoarabinanase	0 (0)
Rhamnogalacturonan I	Rhamnogalacturonase	0 (0)

402

403 **3.4 Chemotaxonomy: metabolite profiling reveals diverse biosynthetic**

404 **capabilities**

405 *3.4.1 Cultivation*

406 Cultivation of *A. burnettii* on agar, liquid and grain-based media resulted in ready production of

407 secondary metabolites (Figure S1). Productivity and metabolite diversity for *A. burnettii* was superior

408 on grains compared with agars and stationary and shaken liquids, with barley superior to wheat,

409 basmati rice and jasmine rice (Figure S2). The methanol extract from *A. burnettii* grown on barley

410 for 21 days was separated by gradient HPLC (Figure 2) and analysed using diode array detection and

411 ESI($\pm$)MS total ion current traces (Figure S33). Assessment of the secondary metabolite diversity

412 was undertaken using UV detection at 210 nm, revealing 58 peaks responsible for the total area under

413 the curve (AUC) from 0.5 to 14.0 min. Analysis of the relative abundance of the secondary

414 metabolites revealed a hyperdispersed distribution, with 20, 30 and 48 peaks accounting for 90%,

415 95% and 99% of the total metabolite AUC, respectively, and the remaining 10 metabolites present at

416 only trace levels.

Metabolite	RT (min)	UV Spectrum $\lambda_{\max}$ nm (Intensity %)	Mass Spectrum (ESI)		
			[M-H] ⁻	[M+H] ⁺	MW
burnettide A	3.39	192 (100), 282s (7), 294 (9)	1144	1146	1145
burnettide B	4.00	194 (100), 282s (7), 294 (9)	972	974	973
burnettide C	4.04	194 (100) 282s (7), 294 (8)	986	988	987
asperlicin E	4.80	206 (100), 228 (89), 270 (19), 280s (18), 312 (10), 324s (8)	420	422	421
(+)-isokotanin C	5.05	212 (100), 296s (38), 313 (58), 322 (62)	409	411	410
(+)-desmethylkotanin	5.29	208 (100), 242s (34), 290s (40), 306 (45), 318s (39)	423	425	424
(+)-isokotanin B	5.63	212 (100), 296s (40), 313 (60), 322 (60)	423	425	424
(+)-isokotanin A	5.88	212 (100), 296s (40), 313 (59), 322 (57)	437	439	436
(+)-kotanin	6.37	208 (100), 294s (51), 306 (58), 318 (46)	-	439	438
ochratoxin A	6.41	192 (100), 218 (54), 248s (16), 334 (17)	402	404	403
burnettram acid A	7.30	192 (100), 226 (34), 282 (60)	769	771	770
14α-hydroxypaspalinine	7.44	192 (67), 230 (100), 274 (26)	448	450	449
hirsutide	7.89	192 (100), 20 (30)	567	569	568
burnettiene A	8.45	192 (23), 218 (24), 318 (63), 334 (100), 352 (96)	537	-	538
anominine	10.16	192 (100), 224 (80), 268s (11), 282 (15)	-	406	405
Paspaline	10.29	192 (64), 230 (100), 280 (25)	420	422	421
emindole SB	10.54	192 (79), 230 (100), 280 (25)	-	406	405

Figure 2. HPLC profile (210 nm) and major metabolites isolated from crude methanol extract of *Aspergillus burnettii* MST-FP2249 after cultivation on barley for 21 days.

3.4.2 Isolation and Characterisation of Metabolites

Seventeen of the 20 metabolites constituting 90% of the total metabolite AUC for *A. burnettii* were purified and formally identified (Figure 2). An initial search of our in-house database of microbial metabolites using COMET (Lacey and Tennant, 2003) tentatively identified the previously reported metabolites ochratoxin A (van der Merwe et al., 1965), anominine (Dowd et al., 1989), (+)-isokotanins A-C (Laakso et al., 1994), (+)-kotanin and (+)-desmethylkotanin (Buechi et al., 1971), and 14α-hydroxypaspalinine (Staub et al., 1993). The identities of these metabolites were confirmed by comparison with known standards and by purification and formal structure elucidation (Tables

430 S1-S11 and Figures S3-S30). The major non-polar metabolites for which standards were unavailable
 431 were purified and analysed by 1D and 2D NMR, revealing previously reported compounds hirsutide
 432 (Lang et al., 2005), paspaline (Fehr and Acklin, 1966), emindole SB (Nozawa et al., 1988), and
 433 asperlicin E (Goetz et al., 1988; Liesch et al., 1988). Two additional non-polar metabolites with highly
 434 diagnostic UV-vis spectra included a poorly soluble tetramic acid, which we recently reported as
 435 burnettramic acid A (Li et al., 2019) and a hitherto unreported polyketide pentaene, burnettiene A,
 436 which will be reported elsewhere. Three additional members of the burnettramic acid family
 437 (burnettramic acids B, A-aglycone and B-aglycone) formed an insoluble layer during solvent
 438 extraction and consequently were not detected in the crude extract. In the polar region, three putative
 439 cyclic peptides, burnettides A-C, were isolated in small quantities and will be reported elsewhere.
 440 Three minor metabolites, eluting at t_R 0.55, 0.72 and 1.70 min, and four unresolved grain lipids,
 441 eluting at 11.06, 11.56, 12.01 and 13.68 min, were not isolated (Figure S33). The structures of the
 442 major secondary metabolites of *A. burnettii* isolated and elucidated are shown in Figure 3.

443 **Figure 3.** Metabolites isolated from *Aspergillus burnettii*

444 Compounds derived from *A. burnettii* cultures are of similar types to *A. alliaceus* and closely related
445 species in *Aspergillus* section *Flavi* (Frisvad et al., 2019). The production of anominine, ochratoxin
446 A, paspalinines, kotanins and isokotanins by *A. burnettii* demonstrate a significant metabolite overlap
447 with *A. alliaceus* (NRRL 315^T = CBS 536.65^T) (Laakso et al., 1994; Nozawa et al., 1994; Junker et
448 al., 2006). It is also relevant that anominine and paspalinine have been reported from *A. nomius* NRRL
449 13137 (Gloer et al., 1989; Staub et al., 1993), while *A. nomius* NRRL 6552 produces anominine and
450 related indole diterpenes (Staub et al., 1992). Interestingly, while paspalinine-related metabolites are
451 present in many related species, it is only in *A. burnettii* that 14 α -hydroxypaspalinine is a dominant
452 metabolite. The production of ochratoxin A by *A. burnettii* represents a further overlap with
453 metabolites from *A. alliaceus* NRRL 315 and NRRL 4181 (Ciegler, 1972; Bayman et al., 2002).
454 Hirsutide is a cyclic tetrapeptide that has been reported as a metabolite of the entomogenous
455 hyphomycete, *Hirsutella* sp. (Lang et al., 2005), but has not previously been recognised as an
456 *Aspergillus* metabolite. *A. burnettii* also produced three additional chemotaxonomic markers: the
457 burnettramic acids, burnettiene A and burnettides A-C.

458

459 The response of *A. burnettii* and its adjacent species *A. alliaceus* to changes in growth media were
460 manifestly different. While both cultures could sustain robust growth and media coverage, *A.*
461 *alliaceus* gave lower levels of secondary metabolite production, even on grains, unless cultivated for
462 long periods (Figure S32). Despite the lower productivity, the results agreed with literature reports
463 with *A. alliaceus* sharing ochratoxin A, (+)-isokotanins A and B, (+)-desmethylokotanin, anominine-
464 related metabolites, and low levels of paspalinine-related metabolites, but not 14 α -
465 hydroxypaspalinine and hirsutide. We can also confirm the previously unrecognised presence of trace
466 levels of burnettramic acids, burnettiene A and burnettide A in *A. alliaceus*. Despite the close overlap
467 of secondary metabolites in both *A. burnettii* and *A. alliaceus*, each species possesses a co-metabolite
468 repertoire unique to that species. In a side by side comparison of the type strains, *A. alliaceus* is a

469 high producer of the kotanin and isokotanin families, while *A. burnettii* is dominated by 14 α -
470 hydroxypaspalinine and hirsutide. In summary, *A. burnettii* is a talented producer of an abundant and
471 diverse array of secondary metabolites, with a chemotaxonomic profile that establishes the species as
472 a distinct member of the *A. alliaceus* clade (Figure S32).

473

474 3.4.3 Bioassay

475 Compounds isolated from *A. burnettii* in sufficient amount were tested on a range of organisms for
476 bioassay, including bacteria (*Bacillus subtilis*, *Staphylococcus aureus*), yeasts (*Candida albicans*,
477 *Saccharomyces cerevisiae*), and mammalian tumour (murine myeloma NS-1) and non-tumour
478 (neonatal foreskin fibroblasts, NFF) cell lines. The majority of the compounds tested displayed
479 antibacterial effects against *B. subtilis* and cytotoxicity against murine myeloma NS-1 cells (Table
480 2). Compounds with notable activity include anominine with antibacterial effect against *B. subtilis*
481 (MIC = 0.4 μ g/mL) and cytotoxicity against NS-1 (MIC = 6.3 μ g/mL) and NFF (MIC = 25 μ g/mL).
482 14 α -Hydroxypaspalinine also displayed antibacterial effect against *B. subtilis* (MIC = 6.3 μ g/mL)
483 and cytotoxicity against NS-1 (MIC = 12.5 μ g/mL) and NFF (MIC = 50 μ g/mL). Burnettramic acid
484 A displayed potent activity against *C. albicans* (MIC = 0.4 μ g/mL) and *S. cerevisiae* (MIC = 0.8
485 μ g/mL), antibacterial activity against *B. subtilis* (MIC = 3.1 μ g/mL) and *S. aureus* (MIC = 6.3
486 μ g/mL), and cytotoxicity against NS-1 (MIC = 25 μ g/mL) and NFF (MIC = 12.5 μ g/mL). Emindole
487 SB displayed potent antibacterial activity against *B. subtilis* (MIC = 0.8 μ g/mL) and potent
488 cytotoxicity against NS-1 (MIC = 0.4 μ g/mL) whereas low cytotoxicity against NFF (MIC = 50
489 μ g/mL).

490

491

Table 2. *In vitro* bioassay of *A. burnettii* metabolites against the bacteria *Bacillus subtilis* (*Bs*) and *Staphylococcus aureus* (*Sa*), the yeasts *Candida albicans* (*Ca*) and *Saccharomyces cerevisiae* (*Sc*), the murine myeloma cell line NS-1 (*Mm*) and non-tumour neonatal foreskin fibroblasts (*Nff*).

Metabolite	MIC (µg/mL)					
	<i>Bs</i>	<i>Sa</i>	<i>Ca</i>	<i>Sc</i>	<i>Mm</i>	<i>Nff</i>
anominine	0.4	>100	>200	>200	6.3	25
14α-hydroxypaspalinine	6.3	>100	>200	>200	12.5	50
(+)-isokotanin B	25	>100	>200	>200	12.5	100
(+)-isokotanin A	12.5	100	>200	>200	25	>100
(+)-kotanin	>100	>100	>200	>200	100	>100
hirustide	25	>100	>200	>200	25	50
ochratoxin A	50	>100	>200	>200	25	>100
paspaline	12.5	>50	>100	>100	50	>50
burnettramycin A	3.1	6.3	0.4	0.8	25	12.5
emindole SB	0.8	>100	>200	>200	0.4	50
Controls	0.2 ^a	3.1 ^a	1.6 ^b	0.4 ^c	0.2 ^d	0.6 ^d

^a ampicillin; ^b nystatin; ^c clotrimazole; ^d sparsomycin

3.5 Draft genome assembly and functional annotation of *A. burnettii*

The enzymatic and metabolic diversity of *A. burnettii* inspired us to sequence the complete genome of the organism to better understand its taxonomy and explore its genetic potential for carbohydrate and secondary metabolism. Based on the frequencies of unique 31-mers in the raw sequencing reads, the genome size of *A. burnettii* was estimated to be 43,158,911 bp. The final assembly consists of 1258 contigs and 1211 scaffolds for a total assembly size of 41,001,558 bp, making up 95% of the estimated genome size. Other assembly quality metrics are listed in Table S22. In total, 12706 gene models were predicted using the funannotate pipeline (Palmer and Stajich, 2019). BUSCO analysis revealed that 95.7% of single copy orthologs from the eurotiomycete dataset are present. In total, 13956 Pfam, 21975 COG/eggNOG, 391 MEROPS, 941 CAZy, 1237 BUSCO, 1034 SignalP, 2899 Phobius and 89019 InterProScan annotations were added. The *A. burnettii* genome has been deposited on GenBank under accession SPNV000000000.

512 Additionally, the previously published *Aspergillus hancockii* genome (FRR 3425) has been re-
513 assembled and annotated and is deposited on GenBank under accession MBFL000000000. Briefly, the
514 updated assembly is made up of 1226 scaffolds for a total size of 39,691,269 bp, compared to 2628
515 scaffolds and 39,903,795 bp total size in the previous assembly (GCA_001696595.1). The maximum
516 scaffold length is 397,609 bp and N50 is 83,265 bp, compared to 165,236 bp and 42,584 bp in the
517 previous assembly, respectively. Further comparison of assembly quality metrics is shown in Table
518 S22. 12583 gene models were predicted and functionally annotated. 91.2% of single copy orthologs
519 from the BUSCO eurotiomycete dataset are present.

520

521 **3.6 Molecular phylogenetic analysis places *A. burnettii* closest to *A. alliaceus*** 522 **within section *Flavi***

523 A phylogenetic tree was inferred from concatenated multiple sequence alignments (MSA) of the
524 internal transcribed spacer (ITS) region, β -tubulin (BenA) and calmodulin (CaM) genes, and the
525 second largest RNA polymerase II subunit (RPB2) (Figure 4 and Table 3). *A. burnettii* is a member
526 of *Aspergillus* subgenus *Circumdati* section *Flavi* and clades closest to *A. alliaceus*. Notably, *A.*
527 *burnettii*, *A. alliaceus*, *A. lanosus*, *A. neoalliaceus* and *A. vanderwermei* form a distinct non-
528 aflatoxigenic group within section *Flavi*, separate to the aflatoxin-producing species *A. flavus* and *A.*
529 *nomius* (Frisvad et al., 2019). This agrees with the initial tentative taxonomic assignment above based
530 on morphological characterisation and metabolic profiling. Furthermore, it mirrors the functional
531 annotation of the CAZyme secretome portfolios, in which *A. burnettii* and *A. alliaceus* are closely
532 aligned, while significant differences are found when comparing to *A. hancockii*.

533

Figure 4. Maximum likelihood tree of *A. burnettii* MST FP2249 (=FRR 5400) inferred from combined internal transcribed spacer (ITS), beta-tubulin (BenA) gene, calmodulin (CaM) gene and RNA polymerase II second largest subunit (RPB2) sequences. Maximum likelihood bootstrap values derived from 1000 bootstrap replicates (left) and posterior probabilities (right) are used as branch support values; asterisks indicate full support. Superscript T denotes ex-type strain. *A. burnettii* clades within subgenus *Circumdati* section *Flavi*, clading closest to *A. alliaceus* CBS 536.65^T. *A. avenaceus* CBS 109.46^T is used as an outgroup.

Table 3. NCBI accessions for marker sequences used in phylogenetic analyses in this study.

Organism	ITS	BenA	CaM	RPB2
<i>Aspergillus burnettii</i>	MK429758	MT211761	MT211762	MT211763
<i>Aspergillus alliaceus</i>	EF661551	EF661465	EF661534	MG517825
<i>Aspergillus avenaceus</i>	AF104446	FJ491481	FJ491496	JN121424
<i>Aspergillus coremiiformis</i>	EF661544	EU014104	EU014112	EU021623
<i>Aspergillus flavus</i>	AF027863	EF661485	EF661508	EF661440
<i>Aspergillus hancockii</i>	KX858342	MT211764	MT211765	MT211766
<i>Aspergillus lanosus</i>	EF661553	EF661468	EF661539	EU021642
<i>Aspergillus leporis</i>	AF104443	EF661499	EF661541	EF661459
<i>Aspergillus magaliesburgensis</i>	MK450649	MK451116	MK451511	MK450802
<i>Aspergillus neoalliaceus</i>	MH279420	MG517763	MG518133	MG517954
<i>Aspergillus nomius</i>	AF027860	AF255067	AY017588	EF661456
<i>Aspergillus togoensis</i>	AJ874113	FJ491477	FJ491489	JN121479
<i>Aspergillus vandermerwei</i>	EF661567	EF661469	EF661540	MG517838

3.7 Biosynthetic gene cluster analysis

3.7.1 *A. burnettii* genome encodes an extensive biosynthetic gene cluster repertoire

547 Analyses of the *A. burnettii* genome revealed that the organism encodes 51 polyketide synthases (PKS,
548 Table S13), of which 25 are highly reducing (HR-PKS), 15 non-reducing (NR-PKS) (including three
549 containing a C-methyltransferase (MT) domain), and 7 partially-reducing (PR-PKS), as well as 3
550 hybrid polyketide synthase-nonribosomal peptide synthetases (PKS-NRPSs) and one truncated PKS
551 homologous to NR-PKSs (Figure 5). The genome also encodes 22 nonribosomal peptide synthetases
552 (NRPS), and 6 NRPS-like synthetases (Table S14), as well as 19 proteins predicted to play a role in
553 terpene biosynthesis (Table S15). When compared to other common species of *Aspergillus*, *A.*
554 *burnettii* can be considered as particularly biosynthetically talented (Figures 6A and B).
555

Figure 5. Synthasase plot of polyketide synthases (PKS) in the *A. burnettii* genome. Synthases are separated into highly-reducing (HR; synthases with enoyl-reductase, keto-reductase and dehydratase domains), partially-reducing (PR) PKS (synthases with at least one reducing domain), non-reducing (NR) PKS (synthases with no reducing domains) and hybrid PKS-NRPS (synthases with both ketosynthase and adenylation domains).

563 Use of the antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) (Blin et al., 2017)
 564 enabled the prediction of 101 BGCs (Table S16) potentially capable of producing a diverse array of
 565 metabolites. Only 18 of these BGCs share any homology to those deposited on the “Minimum
 566 Information about a Biosynthetic Gene Cluster” (MIBiG) database, suggesting *A. burnettii* may
 567 possess unique secondary metabolite production machinery. Use of the ClusterFinder (Cimermanic
 568 et al., 2014) module in antiSMASH resulted in the prediction of an additional 68 potential BGCs, 6
 569 of which sharing homology to BGCs on MIBiG. *A. burnettii* is significantly enriched in BGCs
 570 compared to other model *Aspergillus* species (Table 4). For example, *A. hancockii* and *A. alliaceus*,
 571 which clade closely to *A. burnettii* within *Aspergillus* section *Flavi*, contain 10 and 17 less predicted
 572 BGCs, respectively. Other model Aspergilli differ more substantially, with 21, 37, 56 and 66 less
 573 BGCs in *A. niger*, *A. oryzae*, *A. nidulans* and *A. fumigatus*, respectively.

574

575 **Table 4.** Number of predicted PKS, NRPS, terpene biosynthetic genes and total biosynthetic gene
 576 clusters (BGCs). Number of BGCs are given as those predicted by antiSMASH and the number of
 577 additional BGCs predicted using the ClusterFinder module is shown in parentheses.

Organism	PKS	NRPS	Terpene	Predicted BGCs
<i>Aspergillus burnettii</i>	51	28	19	101 (68)
<i>Aspergillus hancockii</i>	37	26	13	91 (94)
<i>Aspergillus alliaceus</i>	46	22	17	80 (89)
<i>Aspergillus niger</i>	42	26	18	80 (80)
<i>Aspergillus oryzae</i>	30	18	13	70 (95)
<i>Aspergillus nidulans</i>	26	16	9	45 (74)
<i>Aspergillus fumigatus</i>	14	16	8	35 (58)
<i>Aspergillus flavus</i>	30	24	13	69 (117)
<i>Aspergillus parasiticus</i>	24	7	8	65 (98)

578

579 The secondary metabolite repertoire of a species can reflect its taxonomy, especially when comparing
 580 with closely related species. The BCG prediction highlights that *A. burnettii* contains many 'silent'
 581 BGCs not expressed under laboratory cultivation conditions. Tools such as antiSMASH and
 582 Clusterfinder provide an expanded perspective of the metabolic capability of a strain than typically
 583 observed by chemotaxonomic analysis. Indeed, the number of BGCs identified in the genome exceeds

584 the number of metabolites we have isolated from *A. burnettii* in various agar, liquid and grain-based
 585 growth media. The availability of genome sequences allowed us to extend the analysis of secondary
 586 metabolism to all BGCs encoded in each *Aspergillus* genome. As molecular phylogenetic analysis
 587 placed *A. burnettii* within section *Flavi* alongside *A. hancockii*, which we published previously (Pitt
 588 et al., 2017), and is most closely related to *A. alliaceus*, we decided to look more closely at the overlap
 589 in secondary metabolism between these species. Due to the prevalence of important mycotoxins
 590 amongst members of section *Flavi*, we extended this set to include *A. oryzae* 3.042, *A. flavus* AF70
 591 and *A. parasiticus* SU-1, known primarily as producers of aflatoxin, and investigated the presence of
 592 carbohydrate active enzymes (CAZymes). BGCs were first predicted by antiSMASH v4.3; as above,
 593 *A. burnettii* possesses a greater BGC arsenal than other species (Figure 6B, Table 4). For example, *A.*
 594 *burnettii* encodes 15 more PKS BGCs, 2 more NRPS BGCs and 3 more terpene BGCs than *A. flavus*.
 595 This trend is also observed when comparing to *A. oryzae* and *A. parasiticus*.
 596
 597 BGCs were grouped into gene cluster families (GCFs) using the Biosynthetic Genes Similarity
 598 Clustering and Prospecting Engine (BiG-SCAPE) (Navarro-Muñoz et al., 2018). Out of the 1040
 599 BGCs predicted by antiSMASH, 681 GCFs were predicted across the six species, the distribution of
 600 which noticeably mirroring the phylogenetic relationships of each species (Figure 6A). Notably,
 601 though encoding less PKS, NRPS and terpene synthases overall, *A. flavus*, *A. oryzae* and *A.*
 602 *parasiticus* were predicted to contain more GCFs than *A. burnettii*, *A. hancockii* and *A. alliaceus*.
 603 This is likely due to the use of the ClusterFinder module in the antiSMASH prediction, whereby many
 604 putative BGCs are identified that are unrelated to known pathways. For example, *A. flavus* has 69
 605 predicted BGCs satisfying the core prediction rules, and 117 from ClusterFinder; by contrast, *A.*
 606 *burnettii* has 101 and 68, respectively. *A. burnettii* and *A. alliaceus* share 61.31% (103/168) and
 607 59.88% (103/172) of their GCFs, respectively, of which 79 GCFs are exclusive to the two species.
 608 When extended to include *A. hancockii*, the set of shared GCFs drops to just 15; individually, *A.*

609 *burnettii* and *A. hancockii* share only 15.48% (26/168) and 13.9% (26/187) of their GCFs,
 610 respectively (Table S18). Similarly, *A. oryzae* and *A. flavus* share 61.96% (101/163) and 55.19%
 611 (101/183) of their GCFs, with the total of shared GCFs dropping to 50 when *A. parasiticus* is included.
 612 Remarkably, though all species considered here clade within a single section, just 9 GCFs are shared
 613 amongst all of them, suggesting that there is a substantial level of metabolic diversity even among
 614 closely related species. The insights afforded here provide an interesting glimpse at how secondary
 615 metabolite BGCs can serve as sensitive indices of taxonomy.

616

617

618 **Figure 6.** Summary of biosynthetic gene cluster (BGC) and CAZyme repertoire in *A. burnettii* MST
 619 FP2249 and other *Aspergilli*. (A) Dendrogram of shared gene cluster family (GCF) sets in *A.*
 620 *burnettii*, *A. alliaceus*, *A. hancockii*, *A. oryzae*, *A. flavus* and *A. parasiticus* as predicted by BiG-
 621 SCAPE; text in white bubbles at tree nodes indicate the number of GCFs shared by all descendants
 622 of that node. (B) Total polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS) and
 623 terpene-type BGCs in each genome, as predicted by antiSMASH v4.3; each category is inclusive of
 624 hybrid BGCs containing a category-specific synth(et)ase. (C) Absence/presence plot of BGCs
 625 encoding the production of mycotoxins aflatoxin, ochratoxin A, patulin and citrinin. (D) Total genes
 626 mapped to CAZy categories using dbCAN 2.0 (AA: auxiliary activity, CE: carbohydrate esterase,
 627 GT: glycosyltransferase, CB: carbohydrate binding module, GH: glycoside hydrolase, PL:
 628 polysaccharide lyase).

629

630 3.7.2 The search for mycotoxin BGCs confirms *A. burnettii* is incapable of aflatoxin

631 biosynthesis

632 *A. burnettii*, like *A. hancockii*, was originally isolated during a survey for aflatoxin-producing fungi
 633 in areas under peanut cultivation and has since only been found in the same area as its initial isolation.

634 The BGC encoding biosynthesis of aflatoxins is well characterised and is a hallmark feature of section

635 *Flavi*, having been observed in many species of that section (Frisvad et al., 2019). Like *A. alliaceus*,
636 *A. burnettii* was not shown to produce aflatoxins in the various growth substrates and conditions we
637 tested above. Conversely, ochratoxin A is actively produced by *A. burnettii*, and a putative BGC was
638 readily identified in its genome via comparison to the previously reported BGC in *A. carbonarius*
639 ITEM 7444 (Gallo et al., 2012; Gallo et al., 2014; Ferrara et al., 2016). This is typical of species in
640 section *Flavi* which, to current knowledge, all possess either the aflatoxin BGC or ochratoxin A BGC,
641 but never both (Frisvad et al., 2019). Biosynthesis of ochratoxin A requires a PKS to synthesise
642 dihydroisocoumarin (*acOTApks*); a chlorinating enzyme (i.e. halogenase; *acOTAhal*); and a NRPS to
643 attach the phenylalanine moiety to the dihydroisocoumarin ring (*acOTAnrps*). A tBLASTn search
644 against the *A. burnettii* genome using these sequences as a query revealed a BGC located on scaffold
645 152 (Figure 7). Pairwise amino acid alignments with the predicted proteins in this BGC (Table S21)
646 revealed a well-conserved PKS (ETB97_009021; 76% identity to AcOTApks), NRPS
647 (ETB97_009023; 64% identity to AcOTAnrps), halogenase (ETB97_009024; 81% identity to
648 AcOTAhal) and cytochrome P450 (ETB97_009025; 74% identity to AcOTAp450), as well as a
649 putative transcription factor (ETB97_009026; 52% identity to OTA4_ASPC5). Additionally, BGCs
650 homologous to those encoding the production of patulin and citrinin were able to be identified in the
651 *A. burnettii* genome, though production of neither compound was detected in the various growth
652 conditions tested (Figure 6C).

653

654 3.7.3 Linking isolated metabolites to BGCs

655 Like the BGC for ochratoxin biosynthesis above, most of the compounds identified in the culture
656 extracts could be traced back to BGCs using bioinformatics approaches and knowledge derived from
657 prior studies on identical or related compounds using methods detailed previously (Cacho et al., 2015).
658 The asperlicins are a family of peptidyl alkaloids produced by several strains of *A. alliaceus* known
659 to be potent antagonists of the cholecystokinin receptor CCK_A (Chang et al., 1985; Houck et al., 1988).

660 The biosynthetic pathway for asperlicin E has previously been characterised in *A. alliaceus* ATCC
661 20656, and consists of an anthranilate-activating bimodular NRPS (AspA), a FAD-dependent
662 oxygenase (AspB) and a valine-activating monomodular NRPS (AspC) (Haynes et al., 2012). Of
663 these, the only publically available amino acid sequence is for AspB. Searching the AspB sequence
664 against the *A. burnettii* genome returned a hypothetical protein containing a FAD binding domain
665 (ETB97_010954) situated in a predicted BGC on scaffold 91 (Figure 7). ETB97_010954 shares 99.8%
666 identity with AspB and is immediately flanked by a bimodular NRPS (ETB97_010955) and a
667 monomodular NRPS (ETB97_010953), both of which share identical domain architectures with
668 AspA (A-T-C-A-T-C) and AspC (A-T-C), respectively. To further confirm this link, a search was
669 performed against the *A. alliaceus* CBS 536.65 genome (publicly available on the Joint Genome
670 Institute (JGI) Genome Portal). This also resulted in the identification of a FAD binding domain
671 containing protein flanked by bimodular and monomodular NRPSs (Figure 7, Table S20)

672

673 The indole-diterpene paspaline is a key intermediate in the biosynthesis of paxilline and related indole
674 diterpenoids and the genes involved in its biosynthesis have been well-characterised in *Penicillium*
675 *paxilli* (Saikia et al., 2006; Tagami et al., 2013). This pathway consists of a geranylgeranyl
676 diphosphate synthase (PaxG), a prenyltransferase (PaxC) and a FAD-dependent monooxygenase
677 (PaxM) that together synthesise an epoxide-containing indole diterpene precursor, and an indole
678 diterpene cyclase (PaxB) responsible for cyclisation of the precursor to afford the polycyclic scaffold
679 in paspaline. Besides *Penicilli*, *Aspergilli* are also well-known to produce indole diterpene alkaloids.
680 In *A. flavus*, the biosynthesis of paspaline-derived aflatrem is encoded by split gene clusters across
681 two genetic loci, *ATM1* and *ATM2* (Nicholson et al., 2009). Interestingly, in *A. burnettii*, a region
682 highly homologous and syntenic to the *A. flavus ATM1* locus can be identified on scaffold 184 (Figure
683 7), but the corresponding *ATM2* locus is missing. In *A. flavus*, the copy of the indole diterpene cyclase
684 gene (*atmB**) in *ATM1* is mutated and exists as a non-functional relic, while a functional copy of

685 *atmB* is located in *ATM2* (Nicholson et al., 2009). Conversely, a closer examination of the gene cluster
686 on scaffold 184 in *A. burnettii* showed that the *atmB* homolog *ETB97_003558* is likely to be
687 functional. Therefore, together, *ETB97_003556-3558* and *ETB97_12949*, consists of the gene
688 ensemble (functionally analogous to *PaxG/C/M/B*) required for biosynthesis of paspaline and
689 emindole SB (Saikia et al., 2006; Tang et al., 2015), but not the more advanced 14 α -
690 hydroxypaspalinine. Curiously, using the *P. paxilli* *pax* cluster genes as query sequences, we
691 uncovered that *A. burnettii* encodes another set of indole diterpene biosynthetic genes, which
692 correspond to the complete set of genes in the *pax* cluster, except *paxD* (*paxG/A/M/B/C/P/Q*).
693 Unfortunately, due to a fragmented assembly the *A. burnettii* *pax* gene homologs were scattered across
694 multiple small scaffolds, some containing only parts of genes. However, they could all be mapped on
695 to the *P. paxilli* *pax* cluster, with the exception of *paxD* (Figure 7, Table S21). The higher homology
696 of these homologs to the *pax* cluster in *P. paxilli* compared to the *ATM1/ATM2* loci in *A. flavus* suggest
697 that this second cluster has a different origin to *ETB97_003556-3558/12949*. The absence of a
698 *paxD/atmD* homolog, which encodes an aromatic prenyltransferase, correspond to the lack of indole
699 diterpene that contains a prenyl group attached to the indole moiety, as in the case of aflatrem and
700 prenylpaxilline biosynthesis (Nicholson et al., 2009; Scott et al., 2013). Importantly, *A. burnettii*
701 contains the homologs for *paxP/atmP* and *paxQ/atmQ*, which based on previous studies (Nicholson
702 et al., 2009; Scott et al., 2013), are likely involved in the biosynthesis of 14 α -hydroxypaspalinine
703 from paspaline.

704

705 Besides paspaline-related indole diterpenes, several *Aspergilli*, such as *Aspergillus tubingensis*, *A.*
706 *nomius* and *A. flavus*, are also known to concurrently produce other indole diterpenes, such as
707 anominine and aflavinine, which have different polycyclic scaffolds (Frisvad et al., 2014; Frisvad et
708 al., 2019). A more recent study showed that in *A. tubingensis* and *A. flavus*, in addition to the gene
709 clusters required for biosynthesis of paspaline-derived indole diterpenes, an unclustered gene

710 encoding a different indole diterpene cyclase (AtS2B or AfB) was shown to be involved in directing
711 the indole diterpene precursor to biosynthesis of anominine and aflavinine (Tang et al., 2015). Indeed,
712 a gene (*ETB97_001471*) encoding a homolog of the *A. flavus* cyclase AfB can be identified in the *A.*
713 *burnettii* genome (64.29% protein identity). Thus, like in *A. flavus*, the cyclase could direct the indole
714 diterpene precursor biosynthesised by the enzymes encoded in the other clusters toward the
715 biosynthesis of anominine. In the case of *A. burnettii*, there are two complete sets of *atm/paxGCM*
716 homologs in different genetic loci that could biosynthesise the required indole diterpene precursor.
717 These findings on the genetic basis of indole diterpene biosynthesis in *A. burnettii* add an additional
718 layer of complexity to the already known complex evolution of indole diterpene biosynthesis in fungi
719 and highlight the modular and combinatorial nature of BGCs in evolution.

720

721 A series of kotanins/isokotanins were produced by *A. burnettii*. The BGC encoding the bicoumarin
722 kotanin has previously been reported in *A. niger* FGSC A1180 (Gil Girol et al., 2012). A non-reducing
723 PKS (*ktnS*; ANI_1_2226184) synthesises the pentaketidic dihydroxycoumarin backbone, which is
724 then *O*-methylated by an *O*-methyltransferase (*ktnB*; ANI_1_1410184), before an oxidative phenol
725 coupling reaction is catalysed by a cytochrome P450 monooxygenase (*ktnC*; ANI_1_2228184). This
726 gene cluster also includes a methyltransferase (*ktnA*; ANI_1_1408184) and a FAD-binding
727 monooxygenase (*ktnD*, ANI_1_2230184), though it is unclear if these are involved in kotanin
728 biosynthesis. The amino acid sequences of these proteins were retrieved from NCBI and searched
729 against the *A. burnettii* genome. This revealed a BGC located on scaffold 278 consisting of a non-
730 reducing PKS (*ETB97_006286*) that shares the same domain architecture and 62.60% protein identity
731 with *ktnS*, an *O*-methyltransferase (*ETB97_006287*) homologous to *ktnB*, a methyltransferase
732 (*ETB97_006288*) homologous to *ktnA* and a cytochrome P450 (*ETB97_006289*) homologous to *ktnC*
733 (Table S20).

734

735 Hirsutide is an *N*-methylated cyclic tetrapeptide, first isolated from the spider-infecting
736 entomopathogenic fungus *Hirsutella* sp. (Lang et al., 2005). It is similar to the onychocins previously
737 isolated from another Australian fungus, *A. hancockii* FRR 3425, which are *N*-methylated cyclic
738 tetrapeptides possessing two pairs of similar amino acid moieties (L-Phe and L-Val or L-Ile) (Pitt et
739 al., 2017). We previously hypothesised that the onychocins would be produced by either a
740 tetramodular NRPS, following the same co-linearity observed in other NRPS (Schwarzer et al., 2003),
741 or a non-canonical NRPS able to act iteratively. Moreover, our inability to identify any NRPS in the
742 *A. burnettii* genome with a *N*-methyltransferase domain, as present in enniatin synthetase (Haese et
743 al., 1993) or cyclosporin synthetase (Zocher et al., 1986), led us to explore the possibility of a *trans*-
744 acting standalone *N*-methyltransferase. Since then, such a combination has been demonstrated to be
745 required for the biosynthesis of the cyclic peptapeptide cycloaspeptides A1 and E2 in *Penicillium*
746 *soppii* and *Penicillium jamesonlandense* (de Mattos-Shipley et al., 2018). Unfortunately, the
747 sequences in *P. soppii* and *P. jamesonlandense* are unavailable for comparison. A search for NRPS
748 clusters conserved across both *A. burnettii* and *A. hancockii* revealed two homologous pairs of a
749 tetramodular NRPS situated next to a putative methyltransferase. Both NRPSs, ETB97_007099 and
750 BBP40_001644, share identical domain architecture (C-A-T-C-A-T-C-C-A-T-C-A-T-C) and 78%
751 pairwise identities; both *N*-methyltransferases, ETB97_007098 and BBP40_001643, are also highly
752 conserved, sharing 72% identity. Unfortunately, no further identifying information is known for the
753 *Hirsutella* sp. shown to produce hirsutide. Presently, the only publicly available *Hirsutella* genomes
754 on NCBI are *H. minnesotensa*, *H. thompsonii* and *H. rhossiliensis*; a tBLASTn search returned no
755 proteins with any significant homology in any of these genomes. The identity of the two NRPS gene
756 clusters in both *A. burnettii* and *A. hancockii* warrant further investigation.

757

Figure 7. Putative biosynthetic gene clusters (BGCs) for compounds isolated from *A. burnettii* culture extracts and their homology previously characterised BGCs. All genes are drawn to scale; grey boxes linking genes are shaded based on amino acid identity (0% white, 100% black). All gene numbers in *A. burnettii* represent locus tags (prefixed ETB97_). White vertical bars in *A. burnettii* paxilline cluster represent break points where scaffolds were concatenated. Red asterisk indicates a non-functional relic.

3.8 CAZyme repertoire of *A. burnettii* reflects its carbon utilisation profile as well as its taxonomic/phylogenetic position

The CAZyme repertoire of *A. burnettii* was analysed using dbCAN 2.0 (Zhang et al., 2018) and compared to other members of *Aspergillus* section *Flavi* (Figure 6D and Table S23). As in the molecular phylogeny, *A. burnettii* is most like *A. alliaceus*, with the greatest difference observed in the number of auxiliary activity enzymes (3). In comparison to *A. flavus* and *A. hancockii*, *A. burnettii* possesses less of all CAZyme families besides glycosyl transferases (112 compared to 108, 104 and 102 for *A. flavus*, *A. hancockii* and *A. oryzae*, respectively). *A. oryzae* also possesses more of each family besides glycosyl transferases, carbohydrate-binding modules and carbohydrate esterases.

776 Notably, *A. flavus* is particularly rich in glycoside hydrolases, with 304 enzymes identified compared
777 to the 259 found in *A. burnettii*.
778

779 Due to the similarity between the CAZyme repertoires of *A. burnettii* and *A. alliaceus*, a more
780 thorough investigation was performed using the conserved unique peptide pattern (CUPP) method.
781 Peptide-based functional annotation via the CUPP method allows for robust assignment to protein
782 CAZy families and subfamilies as well as prediction of enzyme function of the identified enzyme
783 proteins. Given enzyme function predictions, it is possible to deduce the carbohydrate utilisation
784 capabilities of the species under study. Results of the CUPP analysis of *A. burnettii* is given in Table
785 7. The strongest enzyme function diversity is found among enzymes breaking down cellulose,
786 followed by enzymes modifying their own fungal cell wall and enzymes breaking down pectin and
787 hemicellulose. In addition, several enzyme functions are represented by more than one type of
788 enzymes (belonging to different protein families but having the same enzyme function). This type of
789 profile pattern has been found earlier among strong fungal biomass degraders (Busk et al., 2014).
790 Busk et al. interpreted that such secretome composition, where the assumed most essential enzyme
791 functions are represented by several different types of enzyme proteins, allows the organism to adapt
792 to and cope with varying environmental conditions (e.g. protein stability, enzyme/substrate
793 accessibility, pH and temperature tolerance, varying substrate-recalcitrance, humidity, etc). We also
794 compared the functional annotation of the CAZyme secretome of *A. burnettii* to that of *A. alliaceus*
795 (Table S24). The high similarity of the CAZyme secretome between the two species supports their
796 close phylogeny. In contrast, both similarities and differences are found when comparing the
797 secretome enzyme-function profile of these two species with that of *A. hancockii*. Notably, the types
798 of enzyme function, having duplicate and triplicate types of proteins assigned to the same function,
799 are close to identical among all three species. However, significant differences are found in *A.*
800 *hancockii* with respect to which types of enzyme proteins are assigned to the various functions. As

an interlink between armaments of metabolites and enzyme profiles, it is interesting to note that *A. burnettii* in this study was shown to have an enzyme, tomatinase, originally described to overcome host defence by breaking down plant toxins (Sandrock et al., 1995).

Table 5. Family assignment and predicted function of CAZymes of *A. burnettii* MST FP2249 using the CUPP peptide-based functional annotation method. The selected 80 enzyme CAZyme functions are sorted according to types of complex growth substrates they act on in the majority of reports according to the BRENDA database (Jeske et al., 2019) and CAZypedia (Consortium, 2018). Summarised number of genes for each enzyme function is given as a heat map indicator in the column Number of genes. An extended version of this table including comparison to *A. alliaceus* CBS 536.65 is given in Table S24.

Complex substrate	EC number	Function description	No. genes	Enzyme families representing function		
Starch	3.2.1.20	α -glucosidase	7	4 GH13	3 GH31	-
	3.2.1.3	α -1,4-glucosidase	2	2 GH15	-	-
	3.2.1.10	oligo- α -1,6-glucosidase	4	4 GH13	-	-
	3.2.1.x	maltopentaose-forming α -amylase	2	2 GH13	-	-
Cellulose	3.2.1.21	β -glucosidase	14	2 GH1	11 GH3	1 GH5
	3.1.1.x	cellulose acetate esterase	2	2 CE1	-	-
	3.2.1.4	Endo- β -1,4-glucanase	9	4 GH5	1 GH6	4 GH12
	1.14.99.54	lytic cellulose monooxygenase (C1-hydroxylating)	6	5 AA9	1 AA16	-
	1.14.99.56	lytic cellulose monooxygenase (C4-dehydrogenating)	5	5 AA9	-	-
	3.2.1.176	β -1,4-cellobiosidase (reducing end)	3	3 GH7	-	-
	3.2.1.91	β -1,4-cellobiosidase (non-reducing end)	1	1 GH6	-	-
	1.1.99.18	cellobiose dehydrogenase (acceptor)	1	1 AA3	-	-
	3.1.1.72	Acetylxyylan esterase	3	2 CE3	1 CE5	-
Hemi-cellulose	3.1.1.6	acetylerase	2	2 CE16	-	-
	3.2.1.139	α -glucuronidase	1	1 GH67	-	-
	3.2.1.55	α -N-arabinofuranosidase	7	2 GH43	4 GH51	2 GH62
	3.2.1.8	endo- β -1,4-xylanase	5	3 GH10	2 GH11	-
	3.2.1.37	1,4- β -xylosidase	2	1 GH3	1 GH43	-
	3.2.1.131	xylan α -1,2-glucuronosidase	1	1 GH67	-	-
	3.2.1.151	endo- β -1,4-glucanase	4	4 GH12	-	-
	3.2.1.177	α -D-xyloside xylohydrolase	1	1 GH31	-	-
	3.2.1.x	exo- α -L-1,5-arabinanase	2	2 GH93	-	-
Arabinan	3.2.1.99	endo- α -L-1,5-arabinanase	5	5 GH43	-	-
	4.2.2.10	pectin lyase	7	7 PL1	-	-
Pectin	4.2.2.2	pectate lyase	4	1 PL1	3 PL3	-
	3.2.1.15	polygalacturonase	4	4 GH28	-	-
	3.1.1.11	pectinesterase	3	3 CE8	-	-
	3.2.1.67	α -1,4-galacturonidase	2	2 GH28	-	-
	4.2.2.23	rhamnogalacturonan endolyase	3	3 PL4	-	-
	3.2.1.171	rhamnogalacturonan hydrolase	2	2 GH28	-	-
	3.2.1.40	α -L-rhamnosidase	2	1 GH28	1 GH78	-
	4.2.2.24	rhamnogalacturonan endolyase	1	1 PL26	-	-
	3.2.1.174	rhamnogalacturonan rhamnohydrolase	1	1 GH78	-	-
	3.2.1.x	xylogalacturonan hydrolase	2	2 GH28	-	-
	3.2.1.22	α -galactosidase	2	1 GH27	1 GH36	-
Arabino-galactan	3.2.1.181	endo- β -1,3-galactanase	1	1 GH16	-	-
	3.2.1.145	β -1,3-galactosidase	1	1 GH43	-	-
	3.2.1.89	endo- β -1,4-galactanase	1	1 GH53	-	-
	3.2.1.23	β -galactosidase	5	1 GH2	4 GH35	-
Mannan	3.2.1.25	β -mannosidase	2	2 GH2	0	-

	3.2.1.78	endo- β -1,4-mannosidase	6	2 GH5	1 GH26	3 GH134
Fructan	3.2.1.7	inulinase	3	3 GH32	-	-
	3.2.1.26	β -fructofuranosidase	1	1 GH32	-	-
	3.2.1.80	fructan β -fructosidase	1	1 GH32	-	-
	3.2.1.64	β -2,6-fructan 6-levanbiohydrolase	0	-	-	-
Lignin Fucose Cutin Tomatin	1.10.3.2	laccase	4	4 AA1	-	-
	3.2.1.63	α -L-1,2-fucosidase	1	1 GH95	-	-
	3.1.1.74	cutinase	4	4 CE5	-	-
	3.2.1.x	tomatinase	1	1 GH10	-	-
Fungal polysaccharides	3.2.1.14	chitinase	9	9 GH18	-	-
	1.14.99.53	lytic chitin monooxygenase	5	5 AA11	-	-
	3.5.1.25	<i>N</i> -acetylglucosamine-6-phosphate deacetylase	1	1 CE9	-	-
	3.2.1.17	lysozyme	1	1 GH25	-	-
	3.2.1.132	chitosanase	5	3 GH7	5 GH75	-
	3.2.1.165	exo- β -D-1,4-glucosaminidase	1	1 GH2	-	-
	3.2.1.52	β - <i>N</i> -acetylhexosaminidase	2	1 GH3	1 GH20	-
	3.2.1.11	dextranase	1	1 GH49	-	-
	3.2.1.28	α , α -trehalase	2	1 GH37	1 GH65	-
	3.2.1.113	mannosyl-oligosaccharide α -1,2-mannosidase	5	5 GH47	-	-
	3.2.1.84	α -1,3-glucosidase	2	2 GH31	-	-
	3.2.1.59	endo- α -1,3-glucosidase	8	8 GH71	-	-
	3.2.1.x	exo- α -1,6-mannosidase	1	1 GH125	-	-
	3.2.1.24	α -mannosidase	1	1 GH38	-	-
	3.2.1.x	exo- β -1,3-glucanase (Sun41)	2	2 GH132	-	-
	3.2.1.58	β -1,3-glucosidase	14	8 GH3	2 GH5	4 GH55
	3.2.1.39	endo- β -D-1,3-glucosidase	2	1 GH81	1 GH16	-
	3.2.1.75	endo- β -1,6-glucosidase	1	1 GH5	-	-
	3.2.1.x	exo- β -1,3/1,6- and endo- β -1,4-glucanase	2	2 GH131	-	-
	4.2.2.14	endo- β -1,4-glucuronan lyase	2	1 PL7	1 PL20	-
Glycoproteins	3.2.1.164	galactan endo- β -1,6-galactosidase	1	1 GH5	-	-
	3.2.1.185	β -L-arabinofuranosidase (non-reducing)	1	1 GH142	-	-
	3.2.1.96	endo- β - <i>N</i> -acetylglucosaminidase	2	2 GH18	-	-
	3.2.1.31	β -glucuronidase	2	2 GH154	-	-
	3.2.1.x	d-4,5-unsaturated β -glucuronyl hydrolase	2	2 GH88	-	-
	3.2.1.50	α - <i>N</i> -acetylglucosaminidase	1	1 GH89	-	-
	3.2.1.x	α -1,4-galactosaminogalactan hydrolase	3	3 GH135	-	-
	3.2.1.109	endogalactosaminidase	1	1 GH114	-	-
	3.2.1.x	peptidoglycan hydrolase	1	1 GH18	-	-
	3.2.1.106	mannosyl-oligosaccharide glucosidase	1	1 GH63	-	-

813

814 4. Conclusions

815 Our traditional view of the microbial world as a battleground of microbes with potent chemical
816 arsenals fighting for supremacy has unsurprisingly been shaped by the success of the antibiotic era.
817 As our knowledge of microbial ecology continues to grow, it is now apparent that these interactions
818 are far more nuanced. However, decoding this cryptic language of complex microbial interactions is
819 highly challenging and requires intimate knowledge of the chemistry and biology of the
820 microorganisms. Exploring the interface between taxonomy/phylogeny and (bio)chemical diversity
821 is a powerful approach for elucidation of evolution of the fungal interacting secretome (metabolites

822 and digestive enzymes) as well as for the discovery of new chemical entities, and the integration of
823 genomic analyses further facilitates the navigation of this space. In this study, we have employed
824 taxonomy, genomics, secretome profiling and natural products chemistry to generate a
825 comprehensive chemical and biological profile of *A. burnettii* as an exemplar of such a
826 multidisciplinary approach to microbial biodiscovery. *A. burnettii* is an endemic Australian species
827 that has adapted to an imported agricultural crop (peanuts) and its commensal mycobiota, most
828 notably the pathogen *A. flavus*. The comprehensive data and analyses generated here will serve as a
829 foundation for future exploration of the hidden biosynthetic potential of *A. burnettii* and for further
830 investigation of the complex ecological interactions with peanut-associated Aspergilli. The
831 comparative analysis of both metabolites and enzymes provides a broader understanding of their
832 importance in microbial evolution and as an integrated part of speciation. The relationship between
833 genome, metabolome and secretome compositions provides a more complete picture of the
834 classification and phylogeny of *A. burnettii* as a new recognised species.

835

836

837

838 **Acknowledgements**

839 We acknowledge NRRL for supplying the type strain of *Aspergillus alliaceus*.

840

841 **Funding**

842 This research was funded, in part, by the Australian Research Council (FT160100233, FT130100142)

843 and the Cooperative Research Centres Projects scheme (CRCPFIVE000119).

844

845 **References**

- 846 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012)
847 SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput*
848 *Biol* **5**: 455-477.
- 849 Barrett, K., Jensen, K., Meyer, A.S., Frisvad, J.C., and Lange, L. (2020) Fungal secretome profile
850 categorization of CAZymes by function and family corresponds to fungal phylogeny and taxonomy:
851 Example *Aspergillus* and *Penicillium*. *Sci Rep* **10**: 5158.
- 852 Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R. et al. (2017) UniProt:
853 The universal protein knowledgebase. *Nucleic Acids Res* **45**: D158-D169.
- 854 Bayman, P., Baker, J.L., Doster, M.A., Michailides, T.J., and Mahoney, N.E. (2002) Ochratoxin
855 production by the *Aspergillus ochraceus* group and *Aspergillus alliaceus*. *Appl Environ Microbiol*
856 **68**: 2326-2329.
- 857 Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A. et al. (2017) antiSMASH
858 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids*
859 *Res* **45**: W36-W41.
- 860 Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using
861 DIAMOND. *Nat Methods* **12**: 59-60.
- 862 Buechi, G., Klaubert, D.H., Shank, R.C., Weinreb, S.M., and Wogan, G.N. (1971) Structure and
863 synthesis of kotanin and desmethylkotanin, metabolites of *Aspergillus glaucus*. *J Org Chem* **36**: 1143-
864 1147.
- 865 Bushnell, B. (2014). BBTools software package. URL <https://sourceforge.net/projects/bbmap/>
- 866 Busk, P.K., Lange, M., Pilgaard, B., and Lange, L. (2014) Several genes encoding enzymes with the
867 same activity are necessary for aerobic fungal degradation of cellulose in nature. *PLoS One* **9**:
868 e114138.
- 869 Cairns, T.C., Zheng, X., Zheng, P., Sun, J., and Meyer, V. (2019) Moulding the mold: understanding
870 and reprogramming filamentous fungal growth and morphogenesis for next generation cell factories.
871 *Biotechnol Biofuels* **12**: 77.
- 872 Chang, R., Lotti, V., Monaghan, R., Birnbaum, J., Stapley, E., Goetz, M. et al. (1985) A potent
873 nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from *Aspergillus*
874 *alliaceus*. *Science* **230**: 177-179.
- 875 Chaudhary, N.K., Pitt, J.I., Lacey, E., Crombie, A., Vuong, D., Piggott, A.M., and Karuso, P. (2018)
876 Banksialactones and banksiamarins: isochromanones and isocoumarins from an Australian fungus,
877 *Aspergillus banksianus*. *J Nat Prod* **81**: 1517-1526.
- 878 Chooi, Y.H., and Solomon, P.S. (2014) A chemical ecogenomics approach to understand the roles of
879 secondary metabolites in fungal cereal pathogens. *Front Microbiol* **5**: 640.
- 880 Ciegler, A. (1972) Bioproduction of ochratoxin A and penicillinic acid by members of the *Aspergillus*
881 *ochraceus* group. *Can J Microbiol* **18**: 631-636.
- 882 Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C., Mavrommatis, K. et
883 al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene
884 clusters. *Cell* **158**: 412-421.
- 885 Consortium, C. (2018) Ten years of CAZypedia: a living encyclopedia of carbohydrate-active
886 enzymes. *Glycobiology* **28**: 3-8.
- 887 Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., and Flouri, T. (2019) ModelTest-
888 NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *bioRxiv*:
889 612903.
- 890 Davies, J. (2006) Are antibiotics naturally antibiotics? *J Ind Microbiol Biotechnol* **33**: 496-499.

891 de Mattos-Shipley, K.M.J., Greco, C., Heard, D.M., Hough, G., Mulholland, N.P., Vincent, J.L. et al.
 892 (2018) The cycloaspeptides: uncovering a new model for methylated nonribosomal peptide
 893 biosynthesis. *Chem Sci* **9**: 4109-4117.

894 de Vries, R.P., Riley, R., Wiebenga, A., Aguilar-Orsorio, G., Amillis, S., Uchima, C.A. et al. (2017)
 895 Comparative genomics reveals high biological diversity and specific adaptations in the industrially
 896 and medically important fungal genus *Aspergillus*. *Genome Biol* **18**: 28.

897 Denning, D.W. (2003) Echinocandin antifungal drugs. *The Lancet* **362**: 1142-1151.

898 Dowd, P.F., Wicklow, D.T., and Gloer, J.B. (1989) Nominine, an insecticidal fungal metabolite. In:
 899 United States Dept. of Agriculture, USA., pp. 13 pp. Avail. NTIS Order No. PAT-APPL-17-353 363.

900 Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput.
 901 *Nucleic Acids Res* **32**: 1792-1797.

902 Endo, A. (2004) The origin of the statins. 2004. *Atheroscler Suppl* **5**: 125-130.

903 Fehr, T., and Acklin, W. (1966) Isolation of 2 new indole derivatives from the mycelia of *Claviceps*
 904 *paspali*. *Helv Chim Acta* **49**: 1907-1910.

905 Ferrara, M., Perrone, G., Gambacorta, L., Epifani, F., Solfrizzo, M., and Gallo, A. (2016)
 906 Identification of a halogenase involved in the biosynthesis of ochratoxin A in *Aspergillus*
 907 *carbonarius*. *Appl Environ Microbiol* **82**: 5631-5641.

908 Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L. et al. (2016) The Pfam
 909 protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279-D285.

910 Frisvad, J.C., Petersen, L.M., Lyhne, E.K., and Larsen, T.O. (2014) Formation of sclerotia and
 911 production of indoloterpenes by *Aspergillus niger* and other species in section *Nigri*. *PLoS One* **9**:
 912 e94857.

913 Frisvad, J.C., Hubka, V., Ezekiel, C.N., Hong, S.B., Novakova, A., Chen, A.J. et al. (2019) Taxonomy
 914 of *Aspergillus* section *Flavi* and their production of aflatoxins, ochratoxins and other mycotoxins.
 915 *Stud Mycol* **93**: 1-63.

916 Gallo, A., Knox, B.P., Bruno, K.S., Solfrizzo, M., Baker, S.E., and Perrone, G. (2014) Identification
 917 and characterization of the polyketide synthase involved in ochratoxin A biosynthesis in *Aspergillus*
 918 *carbonarius*. *Int J Food Microbiol* **179**: 10-17.

919 Gallo, A., Bruno, K.S., Solfrizzo, M., Perrone, G., Mule, G., Visconti, A., and Baker, S.E. (2012)
 920 New Insight into the ochratoxin A biosynthetic pathway through deletion of a nonribosomal peptide
 921 synthetase gene in *Aspergillus carbonarius*. *Appl Environ Microbiol* **78**: 8208-8218.

922 Gil Girol, C., Fisch, K.M., Heinekamp, T., Günther, S., Hüttel, W., Piel, J. et al. (2012) Regio- and
 923 stereoselective oxidative phenol coupling in *Aspergillus niger*. *Angew Chem Int Ed* **51**: 9788-9791.

924 Gilchrist, C.L.M. (2020a) synthaser: a Python toolkit for analysing domain architecture of secondary
 925 metabolite megasynthases using NCBI CD-Search. *Zenodo DOI: 105281/zenodo3661588*.

926 Gilchrist, C.L.M. (2020b) cblaster: a Python package for identifying clustered sequence homologs in
 927 NCBI BLAST databases. *Zenodo DOI: 105281/zenodo3660769*.

928 Gloer, J.B., Rinderknecht, B.L., Wicklow, D.T., and Dowd, P.F. (1989) Nominine: a new insecticidal
 929 indole diterpene from the sclerotia of *Aspergillus nomius*. *J Org Chem* **54**: 2530-2532.

930 Goetz, M.A., Monaghan, R.L., Chang, R.S.L., Ondeyka, J., Chen, T.B., and Lotti, V.J. (1988) Novel
 931 cholecystokinin antagonists from *Aspergillus alliaceus*. I. Fermentation, isolation, and biological
 932 properties. *J Antibiot* **41**: 875-877.

933 Haese, A., Schubert, M., Herrmann, M., and Zocher, R. (1993) Molecular characterization of the
 934 enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide
 935 formation in *Fusarium scirpi*. *Mol Microbiol* **7**: 905-914.

936 Haynes, S.W., Gao, X., Tang, Y., and Walsh, C.T. (2012) Assembly of asperlicin peptidyl alkaloids
 937 from anthranilate and tryptophan: a two-enzyme pathway generates heptacyclic scaffold complexity
 938 in asperlicin E. *J Am Chem Soc* **134**: 17444-17447.

939 Houck, D.R., Ondeyka, J., Zink, D.L., Inamine, E., Goetz, M.A., and Hensens, O.D. (1988) On the
940 biosynthesis of asperlicin and the directed biosynthesis of analogs in *Aspergillus alliaceus*. *J Antibiot*
941 **41**: 882-891.

942 Huerta-Cepas, J., Serra, F., and Bork, P. (2016) ETE 3: Reconstruction, analysis, and visualization of
943 phylogenomic data. *Mol Biol Evol* **33**: 1635-1638.

944 Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork,
945 P. (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-
946 Mapper. *Mol Biol Evol* **34**: 2115-2122.

947 Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019) BRENDA in 2019: a
948 European ELIXIR core data resource. *Nucleic Acids Res* **47**: D542-D549.

949 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. et al. (2014) InterProScan 5:
950 genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

951 Junker, B., Walker, A., Cannors, N., Seeley, A., Masurekar, P., and Hesse, M. (2006) Production of
952 indole diterpenes by *Aspergillus alliaceus*. *Biotechnol Bioeng* **95**: 919-937.

953 Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004) A combined transmembrane topology and signal
954 peptide prediction method. *J Mol Biol* **338**: 1027-1036.

955 Kanoh, K., Kohno, S., Asari, T., Harada, T., Katada, J., Muramatsu, M. et al. (1997) (-)-
956 Phenylahistin: A new mammalian cell cycle inhibitor produced by *Aspergillus ustus*. *Bioorg Med*
957 *Chem Lett* **7**: 2847-2852.

958 Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7:
959 improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

960 Kornerup, A., and Wanscher, J.H. (1978) *Methuen Handbook of Colour*. London: Eyre Methuen.

961 Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019) RAXML-NG: a fast,
962 scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**:
963 4453-4455.

964 Laakso, J.A., Narske, E.D., Gloer, J.B., Wicklow, D.T., and Dowd, P.F. (1994) Isokotanins A-C: new
965 bicoumarins from the sclerotia of *Aspergillus alliaceus*. *J Nat Prod* **57**: 128-133.

966 Lacey, E., and Tennant, S. (2003) Secondary metabolites: The focus of biodiscovery and perhaps the
967 key to unlocking new depths in taxonomy. *Microbiol Aust* **24**: 34-35.

968 Lacey, H.J., Vuong, D., Pitt, J.I., Lacey, E., and Piggott, A.M. (2016) Kumbicins A-D: bis-indolyl
969 benzenoids and benzoquinones from an Australian soil fungus, *Aspergillus kumbius*. *Aust J Chem* **69**:
970 152-160.

971 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2017) PartitionFinder 2: New
972 Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological
973 Phylogenetic Analyses. *Mol Biol Evol* **34**: 772-773.

974 Lang, G., Blunt, J.W., Cummings, N.J., Cole, A.L.J., and Munro, M.H.G. (2005) Hirsutide, a cyclic
975 tetrapeptide from a spider-derived entomopathogenic fungus, *Hirsutella* sp. *J Nat Prod* **68**: 1303-
976 1305.

977 Lange, L. (2017) Fungal Enzymes and Yeasts for Conversion of Plant Biomass to Bioenergy and
978 High-Value Products. In *The Fungal Kingdom*: American Society of Microbiology.

979 Lefkove, B., Govindarajan, B., and Arbiser, J.L. (2007) Fumagillin: an anti-infective as a parent
980 molecule for novel angiogenesis inhibitors. *Expert Rev Anti Infect Ther* **5**: 573-579.

981 Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094-3100.

982 Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler
983 transform. *Bioinformatics* **25**: 1754-1760.

984 Li, H., Gilchrist, C.L.M., Lacey, H.J., Crombie, A., Vuong, D., Pitt, J.I. et al. (2019) Discovery and
985 heterologous biosynthesis of the burnettram acids: rare PKS-NRPS-derived bolaamphiphilic
986 pyrrolizidinediones from an Australian fungus, *Aspergillus burnettii*. *Org Lett* **21**: 1287-1291.

987 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence
988 Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

989 Li, Y.F., Tsai, K.J.S., Harvey, C.J.B., Li, J.J., Ary, B.E., Berlew, E.E. et al. (2016) Comprehensive
 990 curation and analysis of fungal biosynthetic gene clusters of published natural products. *Fungal Genet*
 991 *Biol* **89**: 18-28.
 992 Liesch, J.M., Hensens, O.D., Zink, D.L., and Goetz, M.A. (1988) Novel cholecystokinin antagonists
 993 from *Aspergillus alliaceus*. II. Structure determination of asperlicins B, C, D, and E. *J Antibiot* **41**:
 994 878-881.
 995 Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K. et al. (2015)
 996 Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* **11**: 625-631.
 997 Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013) Challenges in homology search:
 998 HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**: e121-e121.
 999 Navarro-Muñoz, J.C., Selem-Mojica, N., Muldowney, M.W., Kautsar, S., Tryon, J.H., Parkinson, E.I.
 1000 et al. (2018) A computational framework for systematic exploration of biosynthetic diversity from
 1001 large-scale genomic data. *bioRxiv*: 445270.
 1002 Nicholson, M.J., Koulman, A., Monahan, B.J., Pritchard, B.L., Payne, G.A., and Scott, B. (2009)
 1003 Identification of two aflatoxin biosynthesis gene loci in *Aspergillus flavus* and metabolic engineering
 1004 of *Penicillium paxilli* to elucidate their function. *Appl Environ Microbiol* **75**: 7469-7481.
 1005 Nozawa, K., Nakajima, S., Kawai, K.-i., and Udagawa, S.-i. (1988) Isolation and structures of
 1006 indoloditerpenes, possible biosynthetic intermediates to the tremorgenic mycotoxin, paxilline, from
 1007 *Emericella striata*. *J Chem Soc Perkin Trans I*: 2607-2610.
 1008 Nozawa, K., Nakajima, S., Kawai, K., Udagawa, S., and Miyaji, M. (1994) Bicomarins from
 1009 ascostromata of *Petromyces alliaceus*. *Phytochemistry* **35**: 1049-1051.
 1010 Palmer, J., and Stajich, J. (2019) nextgenusfs/funannotate: funannotate v1.6.0. *Zenodo*.
 1011 Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal
 1012 peptides from transmembrane regions. *Nat Methods* **8**: 785-786.
 1013 Pierce, N.T., Irber, L., Reiter, T., Brooks, P., and Brown, C.T. (2019) Large-scale sequence
 1014 comparisons with sourmash. *F1000Research* **8**: 1006.
 1015 Piggott, A.M., and Karuso, P. (2004) Quality, not quantity: The role of natural products and chemical
 1016 proteomics in modern drug discovery. *Comb Chem High Throughput Screen* **7**: 607-630.
 1017 Piggott, A.M., and Karuso, P. (2016) Identifying the cellular targets of natural products using T7
 1018 phage display. *Nat Prod Rep* **33**: 626-636.
 1019 Pitt, J., and Hocking, A.D. (2006) Mycotoxins in Australia: biocontrol of aflatoxin in peanuts.
 1020 *Mycopathologia* **162**: 233-243.
 1021 Pitt, J.I., and Hocking, A.D. (2009) *Fungi and Food Spoilage*. New York: Springer.
 1022 Pitt, J.I., Lange, L., Lacey, A.E., Vuong, D., Midgley, D.J., Greenfield, P. et al. (2017) *Aspergillus*
 1023 *hancockii* sp. nov., a biosynthetically talented fungus endemic to southeastern Australian soils. *PLOS*
 1024 *ONE* **12**: e0170254.
 1025 Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018) Posterior
 1026 summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol* **67**: 901-904.
 1027 Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014) MEROPS: the database of
 1028 proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**: D503-D509.
 1029 Saikia, S., Parker, E.J., Koulman, A., and Scott, B. (2006) Four gene products are required for the
 1030 fungal synthesis of the indole-diterpene, paspaline. *FEBS Lett* **580**: 1625-1630.
 1031 Sandrock, R.W., DellaPenna, D., and VanEtten, H.D. (1995) Purification and characterization of beta
 1032 2-tomatinase, an enzyme involved in the degradation of alpha-tomatine and isolation of the gene
 1033 encoding beta 2-tomatinase from *Septoria lycopersici*. *Mol Plant Microbe Interact* **8**: 960-970.
 1034 Schwarzer, D., Finking, R., and Marahiel, M.A. (2003) Nonribosomal peptides: from genes to
 1035 products. *Nat Prod Rep* **20**: 275-287.
 1036 Scott, B., Young, C.A., Saikia, S., McMillan, L.K., Monahan, B.J., Koulman, A. et al. (2013) Deletion
 1037 and gene expression analyses define the paxilline biosynthetic gene cluster in *Penicillium paxilli*.
 1038 *Toxins (Basel)* **5**: 1422-1446.

1039 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015)
 1040 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.
 1041 *Bioinformatics* **31**: 3210-3212.

1042 Smit, A., and Hubley, R. (2015) RepeatModeler Open-1.0 2008-2015 <http://www.repeatmasker.org>.
 1043 Smit, A., Hubley, R., and Green, P. (2015) RepeatMasker Open-4.0 2013-2015
 1044 <http://www.repeatmasker.org>.

1045 Stajich, J., and Palmer, J. (2018) stajichlab/AAFTF: Bugfix Release v0.2.2 (Version v0.2.2). *Zenodo*.
 1046 Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008) Using native and syntenically
 1047 mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637-644.

1048 Staub, G.M., Gloer, K.B., and Gloer, J.B. (1993) New paspalinine derivatives with antiinsectan
 1049 activity from the sclerotia of *Aspergillus nomius*. *Tetrahedron Lett* **34**: 2569-2572.

1050 Staub, G.M., Gloer, J.B., Wicklow, D.T., and Dowd, P.F. (1992) Aspernomine: a cytotoxic
 1051 antiinsectan metabolite with a novel ring system from the sclerotia of *Aspergillus nomius*. *J Am Chem*
 1052 *Soc* **114**: 1015-1017.

1053 Tagami, K., Liu, C., Minami, A., Noike, M., Isaka, T., Fueki, S. et al. (2013) Reconstitution of
 1054 biosynthetic machinery for indole-diterpene paxilline in *Aspergillus oryzae*. *J Am Chem Soc* **135**:
 1055 1260-1263.

1056 Tang, M.C., Lin, H.C., Li, D., Zou, Y., Li, J., Xu, W. et al. (2015) Discovery of unclustered fungal
 1057 indole diterpene biosynthetic pathways through combinatorial pathway reassembly in engineered
 1058 yeast. *J Am Chem Soc* **137**: 13724-13727.

1059 Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008) Gene prediction
 1060 in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* **18**:
 1061 1979-1990.

1062 Thom, C., and Church, M.B. (1926) *The Aspergilli*. Baltimore: The Williams & Wilkins Company.

1063 van der Merwe, K.J., Steyn, P.S., Fourie, L., Scott, D.B., and Theron, J.J. (1965) Ochratoxin A, a
 1064 toxic metabolite produced by *Aspergillus ochraceus*. *Nature* **205**: 1112-1113.

1065 Vesth, T.C., Nybo, J.L., Theobald, S., Frisvad, J.C., Larsen, T.O., Nielsen, K.F. et al. (2018)
 1066 Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section
 1067 *Nigri*. *Nat Genet* **50**: 1688-1695.

1068 Walsh, C., and Tang, Y. (2017) *Natural Product Biosynthesis : Chemical Logic and Enzymatic*
 1069 *Machinery*: Royal Society of Chemistry.

1070 Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z. et al. (2018) dbCAN2: a meta server
 1071 for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95-W101.

1072 Zocher, R., Nihira, T., Paul, E., Madry, N., Peeters, H., Kleinkauf, H., and Keller, U. (1986)
 1073 Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from
 1074 *Tolypocladium inflatum*. *Biochemistry* **25**: 550-553.

1075