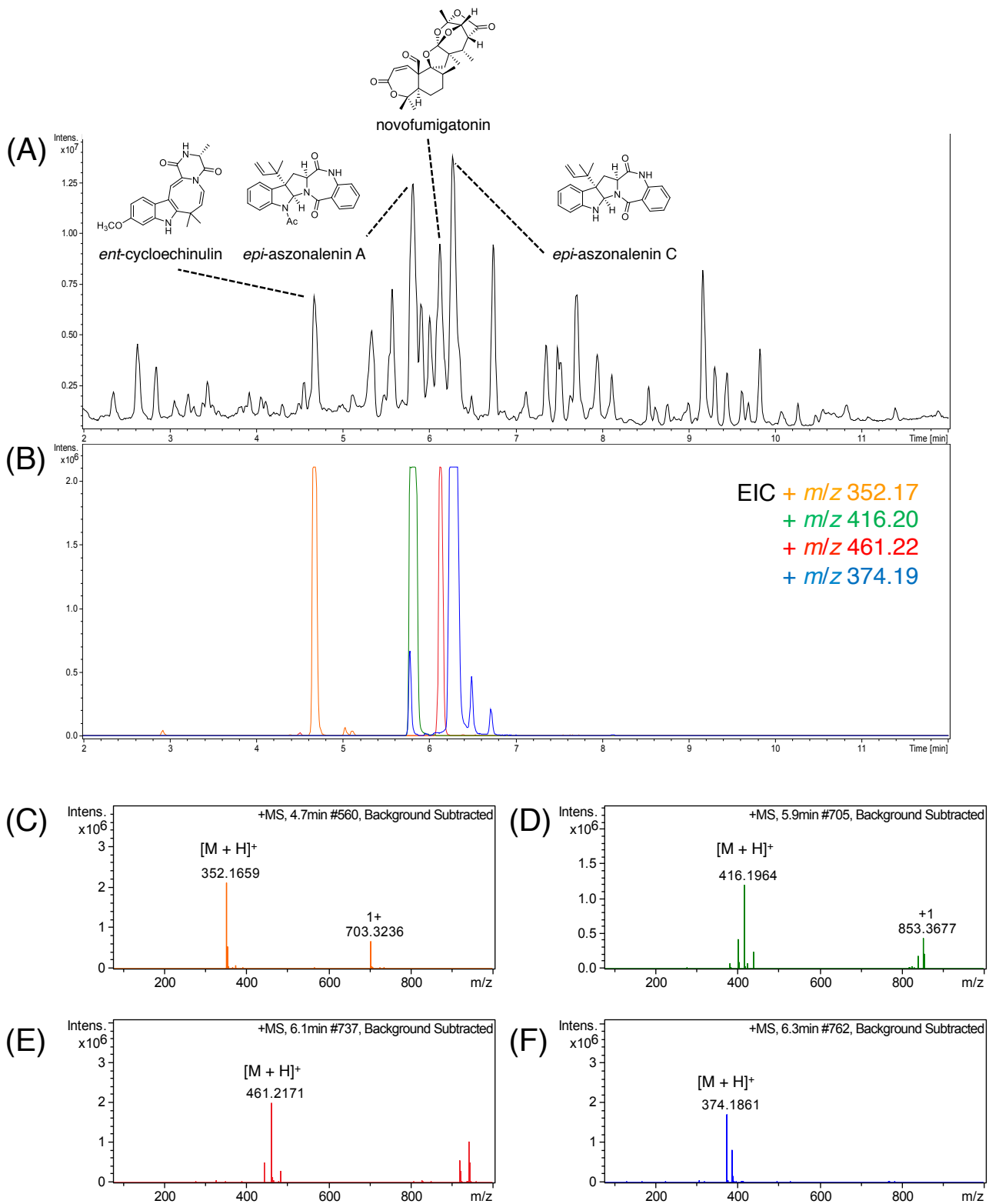
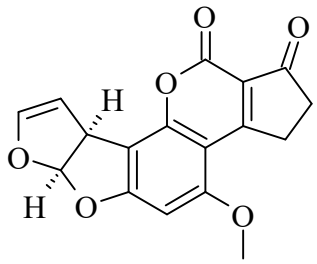
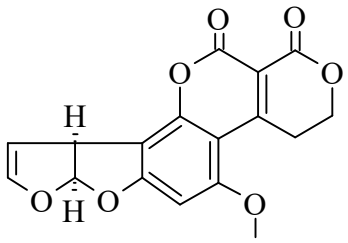


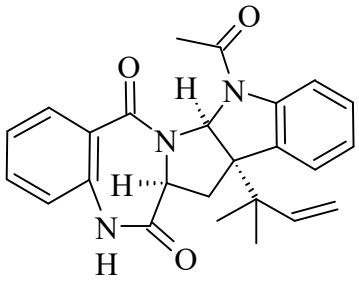
Figure S1 LC-MS analysis of the metabolites from Aspergillus novofumigatus IBT 16806 cultivated on YES agar medium. (A) Total and (B) extracted ion chromatograms, and mass spectra of (C) *ent*-cycloechinulin, (D) *epi*-aszonalenin A, (E) novofumigatonin, and (F) *epi*-aszonalenin A.



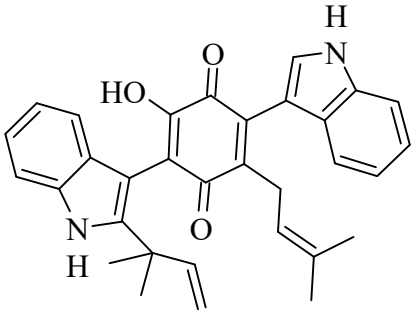
Aflatoxin B1



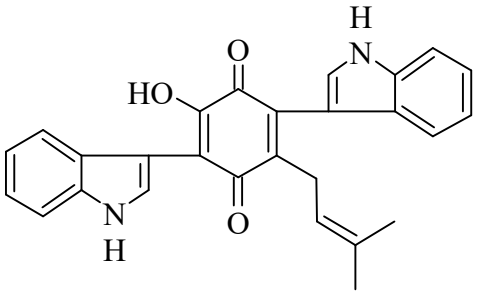
Aflatoxin G1



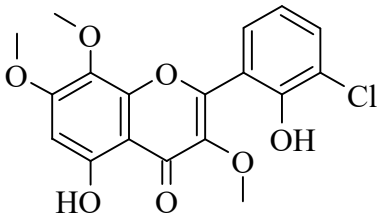
epi-aszonalenin A



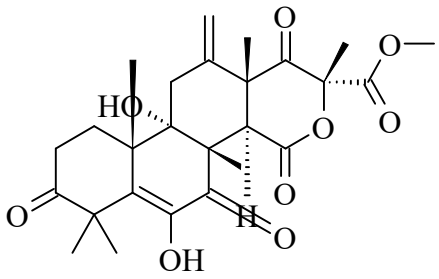
Terrequinone



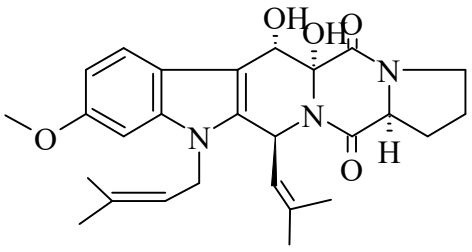
Ochrindole D



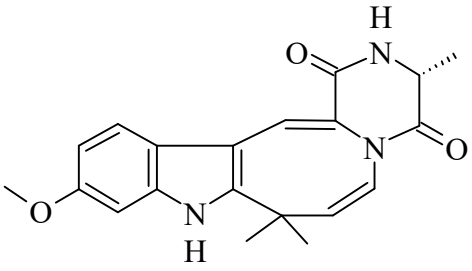
Chlorflavonin



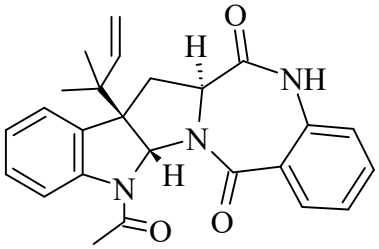
Terretonin



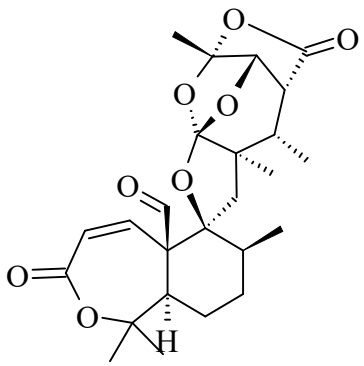
Fumitremorgin B



ent-cycloechinulin



Acetylaszonalenin



Novofumigatonin

Figure S2 Overview of the chemical structures of compounds mentioned in the article (Aflatoxin B1 and G1, *epi*-aszonalenin A, terrequinone, ochrindole D, chlorflavonin, terretonin, fumitremorgin B, *ent*-cycloechinulin, acetylaszonlenin and novofumigatonin).

SI Text

Part 1: Materials and Methods

Strains:

A. campestris (CBS 348.81 = IBT28561, NCBI Taxonomy ID: 1392248) was first isolated in 1979 from soil collected in northern North Dakota. Initially is was placed in the *ochraceus* group (section *Circumdati*) [1]. This was however only based on morphology. In 2000 Rahbæk et al. suggested that *A. campestris* should be placed in section *Candidi* based on chemotaxonomical evidence [2]. In an investigation of section *Candidi* it was further consolidated that *A. campestris* belongs to this section based on among other phylogenetic studies of the calmodulin and β-tubulin genes [3]. *A. campestris* is known to produce a range of interesting chemical compounds including, candidusin C, terphenyllin and chlorflavonin [2, 3].

A. novofumigatus (CBS 117520 = IBT16806, NCBI Taxonomy ID: 1392255) was originally isolated Californian chamise chaparral soil collected in 1965. It is closely related to the pathogenic species *A. fumigatus* and belongs to the section *Fumigati*. It was initially suggested to be a separate species in 2005, this was mainly based on phylogenetic studies of the beta-tubulin, calmodulin and actin genes and on the extrolite profile [4]. *A. novofumigatus* has been recorded in one instance of aspergillosis along with another species in a patient with leukaemia [5].

A. ochraceoroseus (IBT 24754 = CBS 550.77, NCBI Taxonomy ID: 1392256) was first isolated from soil from the Taï national Park in the Ivory Coast in 1978 and based on the morphology it was assigned to ochraceus group (section *Circumdati*) [6]. Looking at the phylogeny of several housekeeping genes and the extrolite profile *A. ochraceoroseus* it is more closely related to subgenus *Nidulantes* and *Veriscolores* [7]. Currently it is believed to be a member of section Ochraceorosei [8, 9].

A. steynii (IBT 23096 = CBS 112812, NCBI Taxonomy ID: 1392250) has been isolated from Arabica green coffee bean from India. It belongs to section *Circumdati* and is closely related to *A. elegans*. *A. steynii* is important in food spoilage since it is known to produce ochratoxin A in addition to several other extrolites [10, 11].

A. candidus (CBS 102.13 = IBT 13984) was isolated in Japan and belongs to section *Candidi* [2, 3].

A. taichungensis (NCBI 482145 =IBT19404, NCBI Taxonomy ID: 482145) was isolated from Soil in Taiwan and belongs to section *Candidi* [2, 3].

The strains used in this study can be seen in Table 1.

Table 1 The species with whole genome sequences used in this study.

Species	Collection number	JGI abbreviation	Reference
A. campestris	CBS 348.81 = IBT28561	Aspcam1	This study
A. novofumigatus	CBS 117520 = IBT16806	Aspnov1	This study
A. ochraceoroseus	CBS 550.77 = IBT 24754	Aspoch1	This study
A. steynii	CBS 112812 = IBT 23096	Aspste1	This study
A. candidus	CBS 102.13 = IBT 13984	Aspcan1	This study
A. taichungensis	NCBI 482145 =IBT19404	Asptaic1	This study
A. oryzae	RIB40	Aspor1	[12, 13]
A. flavus	NRRL3357	Aspfl1	[12]
A. nidulans	FGSC A4	Aspnid1	[12, 14]
A. niger	ATCC 1015	Aspni7	[15]
A. fumigatus Af293	FGSC A1100	Aspfu1	[16]
A. fumigatus A1163	FGSC A1163	Aspfu_A1163_1	[16–19]
N. fischeri/A. fischerianus	NRRL 181	Neofi1	[12, 20]
A. clavatus	NRRL 1	Aspcl1	[12]
A. terreus	NIH2624	Aspte1	[12]
N. crassa	OR74A	Neucr2	[21]
P. chrysogenum		Pench1	These sequence data were produced by the US Department of Energy Joint Genome Institute

			http://www.jgi.doe.gov/ in collaboration with the user community.

Growth of strains and DNA extraction

Biomass for all fungal strains was obtained from shake flasks containing 200ml of complex media CYA [22]. Biomass was isolated by filtering through Miracloth (Millipore, 475855-1R), freeze dried, and stored at -80C. Subsequently, a sample of frozen biomass was used for RNA purification. First hyphae were lysed in a 2ml micro tube, together with a 5mm Steel bead (QIAGEN), placed in liquid nitrogen by using the QIAGEN TissueLyser LT at 45 Hz for 50 seconds. Then the QIAGEN RNeasy mini Plus Kit was used, and the RLT Plus buffer (with 2-mercaptoethanol) was added to the samples, vortexed and spun down. The lysate was then used in step 4 in the instructions provided by the manufacturer, and the protocol was followed from this step. For genomic DNA a protocol inspired by Fulton et al. [23] was used. For details see Additional file 8.

Genome sequencing and assembly. *A. campestris, A. novofumigatus, A. ochraceoroseus* and *A. steynii* were whole-genome sequenced using PacBio RS and assembled using Hierarchical-Based AssembleR (HBAR) which is a developing version derived from Hierarchical Genome Assembly Process (HGAP) [24] allowing larger genomes to be processed.

Unamplified libraries were generated using Pacific Biosciences standard template preparation protocol for creating >10kb libraries. 5ug of gDNA was used to generate each library and the DNA was sheared using Covaris g-Tubes to generate sheared fragments of >10kb in length. A modified version of the protocol was used for 5kb PacBio libraries using a Covaris LE220 focused-ultrasonicator with their Red miniTUBES for DNA shearing. The sheared DNA fragments were then prepared using Pacific Biosciences SMRTbell template preparation kit, where the fragments were treated with DNA damage repair, had their ends repaired so that they were blunt-ended, and 5' phosphorylated. Pacific Biosciences hairpin adapters were then ligated to the

fragments to create the SMRTbell template for sequencing. The SMRTbell templates were then purified using exonuclease treatments and size-selected using AMPure PB beads. Sequencing primer was then annealed to the SMRTbell templates and Version P4 sequencing polymerase was bound to them. The prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences RSII sequencer using Version C2 chemistry and 2 hour sequencing movie run times. Genomes of *A. campestris, A. novofumigatus, A. ochraceoroseus* and *A. steynii* were assembled using HBAR (https://github.com/PacificBiosciences/HBAR-DTK) and polished with quiver.

A. candidus and *A. taichungensis* were whole-genome sequenced using Illumina. 100ng of DNA was sheared to 270bp using the Covaris LE220 (Covaris) and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA biosystems). qPCR was used to determine the concentration of the libraries. Libraries were sequenced on the Illumina Hiseq in 2x150bp format.
The assemblies were produced using combination of Velvet and AllPathsLG v.R47710 assemblers. The raw fastq file was QC filtered to remove artifact/process contamination and then separated into two fastq files, one with mitochondrial data only and the remainder in the target fastq. The target fastq was subsequently assembled using Velvet [25]. The resulting assembly was used to create a long mate-pair library with insert 3000 +/- 90 bp which was then assembled together with the target fastq using AllPathsLG release version R47710, [26].

Transcriptome sequencing and assembly

Stranded cDNA libraries were generated using the Illumina Truseq Stranded RNA LT kits. mRNA was purified from 1µg (100ng for WPOO) of total RNA using magnetic beads containing poly-T oligos. mRNA was fragmented using divalent cations and high temperature. The fragmented RNA was reversed transcribed

using random hexamers and SSII (Invitrogen) followed by second strand synthesis. The fragmented cDNA was treated with end-pair, A-tailing, adapter ligation, and 10 (12 for WPOO) cycles of PCR. qPCR was used to determine the concentration of the libraries. Libraries were sequenced on the Illumina Hiseq. Illumina reads of stranded RNA-seq data were used as input for de novo assembly of RNA contigs. Reads were assembled into consensus sequences using Rnnotator (v. 3.3.1 or later), which consists of three major components: preprocessing of reads, assembly, and postprocessing of contigs [27].

Genome annotation. Annotation of the genomes was completed using the JGI annotation pipeline and made publicly available via JGI fungal genome portal Mycocosm [28, 29].

Also the software tool Secondary Metabolite Unknown Regions Finder (SMURF) was used to predict secondary metabolite gene clusters (SMGC) in the genomes [30].

Sequences were analysed using InterProScan5 in order to investigate potential protein functions [31].

BLASTP. Each protein in each of the genomes has been compared to all other proteins using the BLASTP function from the BLAST+suite version 2.2.27 with a non-restrictive e-value cutoff of 10^{10} [32, 33].

Phylogenetic analysis – CVTree. A phylogenetic tree was constructed using the Composition Vector approach. The web based server CVTree3 was used (http://tlife.fudan.edu.cn/archaea/cvtree/cvtree3/) [34, 35]. The proteome for each of the species were uploaded in fasta format and the K-tuple length was set to 8 and then the project as run resulting in a phylogenetic tree. The tree was visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Identification of species-specific genes

All predicted sets of protein sequences for the 4 genomes in this paper (A. campestris, A. novofumigtus, A.

ochraceoroseus, A. steynii) and the nine reference genomes (*A. nidulans, A. oryzae, A. flavus, A. niger, A. fumigatus Af293, A. fumigatus A1163, N. fischeri, A. clavatus, A.terreus*), and two outgroup species (*N. crassa and P. chrysogenum*) were aligned using the BLASTp function from the BLAST+ suite version 2.2.27 with an e-value <= 10^-10 [32, 33]. These 225 whole-genome BLAST tables were analyzed to identify bi-directional hits in all pairwise comparisons. Using custom Python-scripts, paralogs were first identified within the same genome and grouped into sequence similar families using single linkage, meeting the criterion; the sum of the alignment coverage between the pairwise sequences ≥ 130%, the alignment identity between the pairwise sequences ≥ 50%, and the pairwise hit must be bi-directional (present in both BLAST directions). The orthologs were identified across genomes and grouped into sequence similar families using single linkage meeting the same criterions. Singletons were assigned a family having only one gene member. This allowed for identification of species unique genes. All homologs were assigned functional and structural domains using InterPro version 48 [36, 37] and checked for annotation and sequencing errors by investigating scaffold location and sequence identity.

N6-methyldeoxyadenine analysis

Modification detection was performed using single-molecule real-time (SMRT) sequencing and the PacBio SMRT Analysis 2.3.0 toolkit [38]. Following modification detection, 6mA sites were filtered following Mondo et al., 2017, which includes filtering modifications by both coverage (minimum 15x, maximum determined using the R-boxplot function) and modification quality value (mQV; minimum mQV = 25).

Synteny plots of SMGC – Easyfig. Synteny plots were made using Easyfig version 2.1 (http://easyfig.sourceforge.net) [39]. All GenBank files describing the clusters to be compared were uploaded. The tBLASTx option was used with an e-≤ 0.001 , an alignment length ≥ 50 and an alignment identity ≥ 35. Length of scale was set to 2000 bp under the figure option. If necessary, the sequence can be

reversed under the sub region options. After the options have been set the figure can be created choosing vector format (svg).

Homology of *A. novofumigatus* and *A. fumigatus*

The number of proteins from *A. novofumigatus* with homologs in *A. fumigatus* was determined based on BLASP hit with identity ≥50% and the sum of the query and hit coverage ≥130%.

Synteny of *A. novofumigatus* and *A. fumigatus*

The syneteny between *A. novofumigatus* and *A. fumigatus* was determined using MUMmer (http://mummer.sourceforge.net) [40–42]. 'NUCmer' (NUCleotide MUMmer) was used with standard setting to generate alignments between the two species, followed by 'show-coords' to generate a file with the output from where the total coverage of the genome, maximum block length and mean block length could be calculated based on the *A. novofumigtus* length of the alignments.

Comparison of *A. fumigatus* and *A. novofumigatus* secondary metabolite gene clusters

Comparison of the best hits for *A. fumigatus* and *A. novofumigatus* secondary metabolite gene clusters in other species, based on average percent identity of the backbone proteins. First, BLASTP comparisons of all *A. fumigatus* and *A. novofumigatus* SMGC backbone proteins against all SMGC backbone proteins form the dataset were created. Subsequently, an average of backbone proteins identity was calculated per cluster, the best hit is shown in the heatmap. Backbone proteins are defined as proteins with the annotations PKS(-like), NRPS(-like), hybrid, DMATS and TC.

LC-MS analysis of metabolites from *A. novofumigatus* IBT 16806

Aspergillus novofumigatus IBT 16806 was cultivated on YES agar plate at 25 °C for 7 days, and extracted with ethyl acetate containing 1% formic acid. The extract for LC-MS analysis were injected into a Dionex Ultimate 3000 UHPLC system (Thermo Scientific) - a maXis 3G QTOF orthogonal mass spectrometer (Bruker Daltonics), using Electrospray Ionization with a Kinetex C_{18} column (2.1 i.d. x 100 mm; Phenomenex). Separation was performed with a solvent system of water containing 20 mM formic acid (solvent A) and acetonitrile containing 20 mM formic acid (solvent B), at a flow rate of 0.4 ml/min and a column temperature of 40 °C, using the following program: a linear gradient from 10:90 (solvent B/solvent A) to 100:0 for 10 min, 100:0 for the following 3 min, and a linear gradient from 100:0 to 10:90 within the following 2 min. To illustrate our approach of linking metabolites produced by *A. novofumigatus* to their respective gene clusters, we chose to target our metabolite analysis towards the model compounds novofumigatonin, *ent*-cycloechinulin, *epi*-aszonalenin A and C, since they represent major metabolites produced by this species and because we have them as pure standards in our in-house collection of fungal metabolites [43]. For a full list of metabolites known from *A. novofumigatus* please consult Frisvad & Larsen [44].

References

1. Christensen M. The Aspergillus ochraceus Group: Two New Species from Western Soils and a Synoptic Key. Mycologia. 1982;74:210–25. http://www.jstor.org.proxy.findit.dtu.dk/stable/3792887?origin=crossref&seq=1#page_scan_tab_contents. Accessed 20 Oct 2015.

2. Rahbæk L, Frisvad JC, Christophersen C. An amendment of Aspergillus section Candidi based on chemotaxonomical evidence. Phytochemistry. 2000;53:581–6. doi:10.1016/S0031-9422(99)00596-8.

3. Varga J, Frisvad JC, Samson RA. Polyphasic taxonomy of Aspergillus section Candidi based on molecular, morphological and physiological data. Stud Mycol. 2007;59:75–88.

doi:10.3114/sim.2007.59.10.

4. Hong S-B, Go S-J, Shin H-D, Frisvad JC, Samson RA. Polyphasic Taxonomy of Aspergillus fumigatus and Related Species. Mycologia. 2005;97:1316–29. http://www.jstor.org.proxy.findit.dtu.dk/stable/3762370?seq=1#page_scan_tab_contents. Accessed 20 Oct 2015.

5. Peláez T, Álvarez-Pérez S, Mellado E, Serrano D, Valerio M, Blanco JL, et al. Invasive aspergillosis caused by cryptic Aspergillus species: A report of two consecutive episodes in a patient with leukaemia. J Med Microbiol. 2013;62 PART3:474–8.

6. Bartoli A, Maggi O. Four new species of Aspergillus from Ivory Coast soil. Trans Br Mycol Soc. 1978;71:383–94. doi:10.1016/S0007-1536(78)80064-3.

7. Klich M a, Cary JW, Beltz SB, Bennett C a. Phylogenetic and morphological analysis of Aspergillus ochraceoroseus. Mycologia. 2003;95:1252–60.

8. Cary JW, Harris-Coward PY, Ehrlich KC, Moore GG, Wei Q, Bhatnagar D. Functional and phylogenetic analysis of the Aspergillus ochraceoroseus aflQ (ordA) gene ortholog. Mycologia. 2012;104:857–64.

9. Samson RA, Visagie CM, Houbraken J, Hong S-B, Hubka V, Klaassen CHW, et al. Phylogeny,identification and nomenclature of the genus Aspergillus. Stud Mycol. 2014;78:141–73. doi:10.1016/j.simyco.2014.09.001.

10. Visagie CM, Houbraken J, Frisvad JC, Hong S, Klaassen CHW, Perrone G, et al. Ochratoxin production and taxonomy of the yellow aspergilli (Aspergillus section Circumdati). Stud Mycol. 2014;78:1–61. doi:10.1016/j.simyco.2014.09.001.

11. Frisvad JC, Frank J., Houbraken JAMP, Kuijpers AFA, Samson RA, Frisvad JC. New ochratoxin A

producing species of Aspergillus section Circumdati. Stud Mycol. 2004;50:23–43.

12. Arnaud MB, Cerqueira GC, Inglis DO, Skrzypek MS, Binkley J, Chibucos MC, et al. The Aspergillus Genome Database (AspGD): Recent developments in comprehensive multispecies curation, comparative genomics and community resources. Nucleic Acids Res. 2012;40:653–9.

13. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, et al. Genome sequencing and analysis of Aspergillus oryzae. Nature. 2005;438:1157–61. doi:10.1038/nature04300.

14. Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, Batzoglou S, et al. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature. 2005;438:1105–15. doi:10.1038/nature04341.

15. Andersen MR, Salazar MP, Schaap PJ, Van De Vondervoort PJI, Culley D, Thykaer J, et al. Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res. 2011;21:885–97.

16. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. Nature. 2005;438:1151–6. doi:10.1038/nature04332.

17. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus. PLoS Genet. 2008;4:e1000046. doi:10.1371/journal.pgen.1000046.

18. Ronning CM, Fedorova ND, Bowyer P, Coulson R, Goldman G, Kim HS, et al. Genomics of Aspergillus fumigatus. Rev Iberoam Micol. 2005;22:223–8. http://www.ncbi.nlm.nih.gov/pubmed/16499415. Accessed 10 Dec 2015.

19. Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, et al. Sequencing of

mitochondrial genomes of nine Aspergillus and Penicillium species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics. 2012;13:698. doi:10.1186/1471-2164-13-698.

20. Lonial S, Williams L, Carrum G, Ostrowski M, McCarthy P. Neosartorya fischeri: an invasive fungal pathogen in an allogeneic bone marrow transplant patient. Bone Marrow Transplant. 1997;19:753–5. doi:10.1038/sj.bmt.1700715.

21. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, et al. The genome sequence of the filamentous fungus Neurospora crassa. Nature. 2003;422:859–68. doi:10.1038/nature01554.

22. Samson RA, Houbraken J, Thrane U, Frisvad JC, Andersen B. Food and Indoor Fungi. CBS-KNAW Fungal Biodiversity Centre; 2010. http://findit.dtu.dk/en/catalog/2185758085. Accessed 14 Aug 2017.

23. Fulton TM, Chunwongse J, Tanksley SD. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report. 1995;13:207–9.

24. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10:563–9. doi:10.1038/nmeth.2474.

25. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9. doi:10.1101/gr.074492.107.

26. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8. doi:10.1073/pnas.1017351108.

27. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, et al. Rnnotator : an automated de

novo transcriptome assembly pipeline from stranded RNA-Seq reads Duplicate read removal Multiple Velvet assemblies. 2010.

28. Grigoriev I V., Martinez DA, Salamov AA. Fungal genomic annotation. Appl Mycol Biotechnol. 2006;6 C:123–42.

29. Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014;42:699–704.

30. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010;47:736–41. doi:10.1016/j.fgb.2010.06.003.

31. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. doi:10.1093/bioinformatics/btu031.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. doi:10.1016/S0022-2836(05)80360-2.

33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi:10.1186/1471-2105-10-421.

34. Qi J, Wang B, Hao BI. Whole Proteome Prokaryote Phylogeny Without Sequence Alignment: A K-String Composition Approach. J Mol Evol. 2004;58:1–11.

35. Zuo G, Li Q, Hao B. On K-peptide length in composition vector phylogeny of prokaryotes. Comput Biol Chem. 2014;53:166–73. doi:10.1016/j.compbiolchem.2014.08.021.

36. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. Nucleic Acids Res. 2009;37 SUPPL. 1:211–5.

37. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2014;43:D213-221. doi:10.1093/nar/gku1243.

38. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7:461–5. doi:10.1038/nmeth.1459.

39. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–10. doi:10.1093/bioinformatics/btr039.

40. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

41. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 2002;30:2478–83. doi:10.1093/nar/30.11.2478.

42. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. Alignment of whole genomes. Nucleic Acids Res. 1999;27. http://mummer.sourceforge.net/MUMmer.pdf. Accessed 24 Nov 2017.

43. Nielsen KF, Månsson M, Rank C, Frisvad JC, Larsen TO. Dereplication of Microbial Natural Products by LC-DAD-TOFMS. J Nat Prod. 2011;74:2338–48. doi:10.1021/np200254t.

44. Frisvad JC, Larsen TO. Extrolites of Aspergillus fumigatus and Other Pathogenic Species in Aspergillus Section Fumigati. Front Microbiol. 2016;6 January:1–14.

Part 2: Protocol for preparation of Fungal DNA

The protocol described below has successfully been employed to isolate genome-sequencing grade genomic DNA for more than 200 different *Aspergillus* species.

List of Materials:

D-Sorbitol (Sigma, S1876 – CAS 50-70-4)
Tris-Base (Sigma 7-9, T1378 – CAS 7786-1)
37% HCl (Th. Geyer, 836,1000)
EDTA (Merck, 324503 – CAS 6381-92-6)
Sodium Cloride (NaCl) (AppliChem A1371,9010 – CAS 7647-14-5]
Cetyl trimethylammonium bromide (CTAB) (Sigma 52365 – CAS 57-09-0)
Sarkosyl NL (Sigma, L5777 – CAS 137-16-6)
Polyvinylpyrrolidone (PVP) (Sigma PVP-40T – CAS 9003-39-8)
Proteinase K (NEB P8107S)
Potassium acetate (J. T. Baker 0129910025 – CAS 127-08-2)
Phenol:Chloroform:Isoamylalcohol (25:24:1) (Sigma P3803)
Sodium acetate (J. T. Baker 9914011001 – CAS 6131-90-4]
96 % Ethanol (vwr chemicals)
70 % Ethanol (vwr chemicals)
Isopropanol (Merck, 109634 – CAS 67-63-0)
Liquid nitrogen
Sodium Hydroxide (Sigma S5881 – CAS 1310-73-2)
RNase A (Sigma R-4875 – CAS 9001-99-4)

Preparation of liquid Media

Buffers:

- 5M Potassium acetate (pH 7.5): 122.5 g potassium acetate and ddH_2O up to 250 mL. pH adjusted with acetic acid.
- 3M Sodium acetate: 81.65 g sodium acetate and ddH_2O up to 200 mL.
- 1% PVP: 2 g PVP in 200 mL ddH_2O
- 5% Sarkosyl: 10 g Sarkosyl in 200 mL ddH_2O.
- 1M Tris-HCl (pH 9): 60.57 g Tris-base and 4.81 ml 37% HCl. Add ddH_2O up to 500 mL.
- 0.5M EDTA: 116.4 g EDTA. Add ddH_2O up to 500 mL. Add Sodium Hydroxide pellet until pH reach 8.0.
- Buffer A: 31.9 g Sorbitol, 50 mL 1M Tris-HCl (pH 9), 5 mL 0.5M EDTA (pH 8) and ddH_2O up to 500 mL.
- Buffer B: 100 mL 1M Tris-HCl (pH 9), 50 mL 0.5M EDTA, 58.44 g NaCl, 10 g CTAB. Add ddH_2O up to 500 mL.
- TE (pH 9): 1.21 g Tris-base, 0.37 g EDTA. Add ddH_2O up to 1000 mL.

All solutions above are to be autoclaved!

- Lysis Buffer: For 10 ml pr. sample use: 3.75 ml Buffer A; 3.75 ml Buffer B; 1.5 ml 5 % Sarkosyl; 1 ml 1 % PVP; 100 µl Proteinase K
- RNase A: Dissolve 10 mg dry powder in 1 ml ddH_2O

Equipment:

- Nanodrop: NanoDrop ND 1000 Spectrophotometer or NanoDrop Lite from Qiagen.
- Qubit: Qubit 1.0 fluorometer from Invitrogen and Qubit dsDNA BR Assay Kit (Q32853) from ThermoFisher.
- Mortar and pestle.
- Centrifuge for 50ml Falkon tubes at 4°C.

Protocol:

1. Pre-heat Buffer B at 65°C
2. Prepare Lysis Buffer just before use and keep at 65°C.
3. Transfer freeze-dried mycelia into a mortar and cover with liquid nitrogen. Grind material and transfer to a 50 ml Falcon tube as soon as all liquid nitrogen has evaporated. Powder in the tube should not exceed the 5ml mark, but a minimum of 3 ml is recommended. Note powder must not thaw.
4. Add 10 ml Lysis Buffer and mix vigorously by vortexing.
5. Incubate for 30 min at 65°C. Mix frequently by inverting the tube.
6. Add 3.35 ml 5 M Potassium acetate. Mix gently by inverting the tube 5-7 times. Incubate solution 30 min on ice.
7. Centrifuge for 30 min at 5,000 g at 4°C.
8. Transfer the supernatant, approximately 9mL, to a new 50 ml Falcon tube and add 5ml of Phenol:Chloroform:Isoamylalcohol (25:24:1). Mix gently 5-7 times.
9. Centrifuge 20 min at 4,000 g at 4 °C.
10. Transfer the aqueous phase (~8mL) to a new 50 ml Falcon tube. Note avoid any material from the interphase.
11. Add 100 µl RNase A (10 mg/ml) and mix gently. Incubate at room temperature for 30-60 min.
12. Add 1/10 volume of 3M Sodium acetate and 1 volume ice-cold 96% Ethanol (Alternatively, Isopropanol can be used, but it may adversely influence A260/A280 measurements). Incubate solution at 20°C for 30 min.
13. Centrifuge for 30 min at 10,000 g and 4°C
14. Discard the supernatant.
15. Wash the pellet with 2 ml 70 % ethanol and pipette as much away without disturbing the pellet.

16. Dry the pellet at room temperature until all ethanol has evaporated (approximately 15 minutes).
    Note: do not let the pellet dry out!
17. Dissolve the pellet in 500 µL TE. This may take over-night incubation at room temperature with light shaking. Transfer DNA solution to a 2 ml Eppendorf tube.
18. Take a sample for DNA quality assessments (see below) and store the remaining DNA solution at -20 ^{o}C until further use.
19. For testing DNA quality:
    Make a 20-fold dilution of the DNA solution (from step 18) in a 1.5 ml Eppendorf tube to a total volume of 100 µl.
    A. Run a 5-10 µl diluted sample on an agarose gel to estimate the quality and concentration.
    B. Use the nanodrop for A_{260}/A_{280} measurements. Ratios should be in the range of 1.6-2.2.
    C. Use the Qubit to determine DNA concentration estimations. Good predations fall in the range of 20-200 ng/µl DNA in stock solution.

Table S1 Overview of the most common InterPro domains for the unique genes in the investigated species and the number of times the InterPro domain is found in a unique gene in the investigated species.

Species	Total predicted proteins	Total unique proteins	Unique genes %	InterPro annotated unique	Most common IPR, # of IPR in unique genes						
					IPR001138	IPR016040	IPR002110	IPR016196	IPR011009	IPR007219	IPR001128
A. campestris	9764	2162	22%	670 (31%)	43	34	35	21	19	10	28
A. clavatus	9121	1053	12%	349 (33%)	20	23	18	20	6	7	21
A. flavus	12604	1953	15%	659 (34%)	52	46	32	48	25	37	22
A. fumigatus Af293	9781	188	2%	67 (36%)	0	2	5	3	1	0	1
A. fumigatus A1163	9916	343	3%	100 (29%)	8	1	8	5	7	3	3
A. niger ATCC 1015	11910	3168	27%	1232 (39%)	135	94	63	71	37	75	47
A. nidulans	10680	2391	22%	943 (39%)	103	63	37	46	17	47	36
A. novofumigatus	11549	1695	15%	462 (27%)	26	37	30	20	14	13	36
A. ochraceoroseus	8924	1881	21%	519 (28%)	57	26	6	12	27	20	18
A. oryzae	12031	1842	15%	635 (34%)	36	26	46	32	22	28	16
A. steynii	13211	3520	27%	1270 (36%)	163	114	67	54	27	72	43
A. terreus	10406	2117	20%	905 (43%)	55	66	54	45	37	52	37
A. fischerianus	10406	704	7%	249 (35%)	6	21	23	14	11	6	9
N. Crassa	10785	7063	65%	3471 (49%)	147	91	53	66	81	68	28
P. Chrysogenum		3215		1101 (34%)	126	64	44	35	57	52	30

IPR ID	Description
IPR001138	Fungal transcriptional regulatory protein, N-terminal
IPR016040	NAD(P)-binding
IPR002110	Ankyrin
IPR016196	MFS general substrate transporter
IPR011009	Protein kinase-like
IPR007219	Fungal specific transcription factor
IPR001128	Cytochrome P450

Top	IPR with most counts
Second	IPR with second most counts
Third	IPR with third most counts

Table S2 Allergens from *A. novofumigatus.* A list of allergens from *A. fumigatus* (from www.allergome.org) and the orthologs in *A. novofumigatus* including the percent identity of the BLAST comparison.

Allergen name	***A. fumigatus* AF293 accession**	***A. novofumigatus* orthologue**	**% ID**
Asp_f1	AFUA_5G02330	jgi-Aspnov1-365359-e_gw1.2.4275.1	98.86
Asp_f2	AFUA_4G09580	jgi-Aspnov1-432190-fgenesh1_pg.5_#_493	93.56
Asp_f3	AFUA_6G02280	jgi-Aspnov1-388041-estExt_Genewise1Plus.C_3_t50474	97.62
Asp_f4	AFUA_2G03830	jgi-Aspnov1-432819-fgenesh1_pg.6_#_39	94
Asp_f5	AFUA_8G07080	jgi-Aspnov1-395206-estExt_Genewise1Plus.C_7_t30340	95.58
Asp_f6	AFUA_1G14550	jgi-Aspnov1-512129-estExt_fgenesh1_pm.C_4_t10311	99.47
Asp_f7	AFUA_4G06670	jgi-Aspnov1-443000-fgenesh1_kg.5_#_844_#_Locus964v1rpkm138.38	92.6
Asp_f8	AFUA_2G10100	jgi-Aspnov1-30553-CE30552_16949	90
Asp_f9	AFUA_1G16190	jgi-Aspnov1-430704-fgenesh1_pg.4_#_185	94.07
Asp_f10	AFUA_5G13300	jgi-Aspnov1-363785-e_gw1.2.2138.1	95.19
Asp_f11	AFUA_2G03720	jgi-Aspnov1-426603-fgenesh1_pg.1_#_378	93.57
Asp_f12	AFUA_5G04170	jgi-Aspnov1-509883-estExt_fgenesh1_pm.C_2_t10392	99.27
Asp_f13	AFUA_4G11800	jgi-Aspnov1-431992-fgenesh1_pg.5_#_295	94.29
Asp_f15	AFUA_2G12630	jgi-Aspnov1-430704-fgenesh1_pg.4_#_185	94.07
Asp_f17	AFUA_4G03240	jgi-Aspnov1-409635-estExt_Genewise1.C_5_t50054	91.33
Asp_f18	AFUA_4G11800	jgi-Aspnov1-431992-fgenesh1_pg.5_#_295	94.3
Asp_f22	AFUA_6G06770	jgi-Aspnov1-367285-e_gw1.3.3588.1	99.09
Asp_f23	AFUA_2G11850	jgi-Aspnov1-447552-estExt_Genemark1.C_1_t30040	94.23
Asp_f26	AFUA_1G06830	jgi-Aspnov1-370658-e_gw1.4.1124.1	86.49
Asp_f27	AFUA_3G07430	jgi-Aspnov1-453038-estExt_Genemark1.C_6_t10423	90.8
Asp_f28	AFUA_6G10300	jgi-Aspnov1-367561-e_gw1.3.3658.1	96.1
Asp_f29	AFUA_5G11320	jgi-Aspnov1-445073-fgenesh1_kg.8_#_93_#_Locus606v1rpkm225.43	74.26
Asp_AfCalAp	AFUA_3G09690	jgi-Aspnov1-515921-estExt_fgenesh1_pm.C_100047	94.79
Asp_f_chitosanase	AFUA_4G01290	jgi-Aspnov1-412759-estExt_Genewise1.C_8_t10196	94.21
Asp_f_AT	AFUA_1G09470	jgi-Aspnov1-512574-estExt_fgenesh1_pm.C_4_t20289	91.19
Asp_f_catalase	AFUA_3G02270	jgi-Aspnov1-440530-fgenesh1_kg.3_#_979_#_Locus6690v1rpkm8.88	50.21
Asp_f_DPPV	AFUA_2G09030	jgi-Aspnov1-453870-estExt_Genemark1.C_7_t10446	88.57
Asp_f_glucosidase	AFUA_1G05770	jgi-Aspnov1-431600-fgenesh1_pg.4_#_1081	93.36
Asp_f_GST	AFUA_6G09690	jgi-Aspnov1-458931-fgenesh1_pm.3_#_489	96.67
Asp_f_GT	AFUA_6G11390	jgi-Aspnov1-398517-estExt_Genewise1.C_1_t10227	88.44
Asp_f_IAO	AFUA_6G03620	jgi-Aspnov1-413975-estExt_Genewise1.C_9_t20169	51.01
Asp_f_IPMI	AFUA_2G11260	jgi-Aspnov1-500519-estExt_fgenesh1_pg.C_1_t20448	96.27
Asp_f_LPL1	AFUA_4G08720	jgi-Aspnov1-516331-estExt_fgenesh1_pm.C_120063	94.96
Asp_f_LPL3	AFUA_3G14680	jgi-Aspnov1-516331-estExt_fgenesh1_pm.C_120063	90.83
Asp_f_mannosidase	AFUA_1G14560	jgi-Aspnov1-460012-fgenesh1_pm.4_#_314	94.65
Asp_f_MDH	AFUA_7G05740	jgi-Aspnov1-462836-fgenesh1_pm.7_#_121	96.19
Asp_f_PL	AFUA_2G00760	jgi-Aspnov1-499715-estExt_fgenesh1_pg.C_1_t10092	92.83
Asp_f_PUP	AFUA_5G03520	jgi-Aspnov1-438430-fgenesh1_kg.2_#_311_#_Locus9805v1rpkm2.72	95.05
Asp_f_SXR	AFUA_2G15430	jgi-Aspnov1-47013-CE47012_27880	99.57
Asp_f_CP	AFUA_8G01670	jgi-Aspnov1-445108-fgenesh1_kg.8_#_128_#_Locus912v1rpkm147.61	96.31
Asp_f_FDH	AFUA_6G04920	jgi-Aspnov1-440417-fgenesh1_kg.3_#_866_#_Locus1540v1rpkm84.43	95.16

Table S3 Virulence factors from *A. fumigatus*. A collection of virulence factors known from *A. fumigatus* and the best orthologs *in A. novofumigatus* along with the percent identity of the BLAST comparison.

***A. fumigatus* AF293 accession**	**Common name**	**Gene function**	***A. novofumigatus* orthologue**	**% ID**
AFUA_1G01550	zrfA	High affinity zinc ion transporter, putative [Source:UniProtKB/TrEMBL;Acc:Q4WKR5]	jgi-Aspnov1-378097-e_gw1.9.842.1	91.92
AFUA_1G05800	mkk2	MAP kinase kinase (Mkk2), putative [Source:UniProtKB/TrEMBL;Acc:Q4WJJ0]	jgi-Aspnov1-451544-estExt_Genemark1.C_4_t30113	90.68
AFUA_1G09280	ptcB	Protein phosphatase 2C, putative [Source:UniProtKB/TrEMBL;Acc:Q4WTH5]	jgi-Aspnov1-460512-fgenesh1_pm.4_#_814	91.44
AFUA_1G10880	pmcA	P-type calcium ATPase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WT17]	jgi-Aspnov1-420834-gm1.5181_g	92.44
AFUA_1G14660	laeA	Regulator of secondary metabolism LaeA [Source:UniProtKB/TrEMBL;Acc:Q4WRY5]	jgi-Aspnov1-370504-e_gw1.4.2095.1	98.66
AFUA_1G15440	ags3	Alpha-1,3-glucan synthase Ags3 [Source:UniProtKB/TrEMBL;Acc:Q4WRQ8]	jgi-Aspnov1-447502-estExt_Genemark1.C_1_t20487	54.83
AFUA_1G16950	pig-a	Phosphatidylinositol:UDP-GlcNAc transferase subunit PIG-A [Source:UniProtKB/TrEMBL;Acc:Q4WRA7]	jgi-Aspnov1-405712-estExt_Genewise1.C_4_t10229	97.96
AFUA_2G01260	srbA	HLH transcription factor, putative [Source:UniProtKB/TrEMBL;Acc:Q4WIN1]	jgi-Aspnov1-455802-fgenesh1_pm.1_#_129	95.14
AFUA_2G07680	sidA	L-ornithine N5-oxygenase SidA [Source:UniProtKB/TrEMBL;Acc:E9QYP0]	jgi-Aspnov1-447204-estExt_Genemark1.C_1_t20174	96.01
AFUA_2G07770	rasB	Ras small monomeric GTPase RasB [Source:UniProtKB/TrEMBL;Acc:Q4X241]	jgi-Aspnov1-363087-e_gw1.1.2643.1	99.16
AFUA_2G08360	pyrG	Orotidine 5'-phosphate decarboxylase [Source:UniProtKB/Swiss-Prot;Acc:O13410]	jgi-Aspnov1-399623-estExt_Genewise1.C_1_t30405	97.81
AFUA_2G11270	ags2	Alpha-1,3-glucan synthase Ags2 [Source:UniProtKB/TrEMBL;Acc:Q4X143]	jgi-Aspnov1-447502-estExt_Genemark1.C_1_t20487	96.78
AFUA_2G12200	pkaC1	cAMP-dependent protein kinase catalytic subunit PkaC1 [Source:UniProtKB/TrEMBL;Acc:Q4X0V1]	jgi-Aspnov1-363483-e_gw1.1.1181.1	90.24
AFUA_2G12640	gprD	Integral membrane protein [Source:UniProtKB/TrEMBL;Acc:Q4X0Q7]	jgi-Aspnov1-456628-fgenesh1_pm.1_#_955	90.43
AFUA_2G17530	Melanin cluster, arb2	Conidial pigment biosynthesis oxidase Arb2 [Source:UniProtKB/TrEMBL;Acc:E9RBR0]	jgi-Aspnov1-501077-estExt_fgenesh1_pg.C_1_t40073	92.15
AFUA_2G17540	Melanin cluster, abr1	Conidial pigment biosynthesis oxidase Abr1/brown 1 [Source:UniProtKB/TrEMBL;Acc:Q4WZB4]	jgi-Aspnov1-417292-gm1.1639_g	88.61
AFUA_2G17550	Melanin cluster, ayg1	Conidial pigment biosynthesis protein Ayg1 [Source:UniProtKB/TrEMBL;Acc:Q4WZB3]	jgi-Aspnov1-438093-fgenesh1_kg.1_#_1767_#_Locus2445v1rpkm48.63	94.2
AFUA_2G17560	Melanin cluster, arp2	Conidial pigment biosynthesis 1,3,6,8-tetrahydroxynaphthalene reductase Arp2 [Source:UniProtKB/TrEMBL;Acc:E9QUT3]	jgi-Aspnov1-460884-fgenesh1_pm.5_#_30	49.22
AFUA_2G17580	Melanin cluster, arp1	Probable scytalone dehydratase [Source:UniProtKB/Swiss-Prot;Acc:O14434]	jgi-Aspnov1-457049-fgenesh1_pm.1_#_1376	95.15
AFUA_2G17600	Melanin cluster, alb1	Conidial pigment polyketide synthase PksP/Alb1 [Source:UniProtKB/TrEMBL;Acc:Q4WZA8]	jgi-Aspnov1-448094-estExt_Genemark1.C_1_t40143	94.78
AFUA_3G05650	orlA	Alpha,alpha-trehalose-phosphate synthase subunit Tps2, putative [Source:UniProtKB/TrEMBL;Acc:Q4WWF5]	jgi-Aspnov1-514040-estExt_fgenesh1_pm.C_6_t10212	97.88
AFUA_3G09820	dvrA	C2H2 transcription factor, putative [Source:UniProtKB/TrEMBL;Acc:Q4WXK4]	jgi-Aspnov1-462495-fgenesh1_pm.6_#_581	93.38
AFUA_3G11250	ace2	C2H2 transcription factor (Swi5), putative [Source:UniProtKB/TrEMBL;Acc:Q4WXZ7]	jgi-Aspnov1-423283-gm1.7630_g	87.45
AFUA_3G11970	pacC	pH-response transcription factor pacC/RIM101 [Source:UniProtKB/Swiss-Prot;Acc:Q4WY67]	jgi-Aspnov1-514555-estExt_fgenesh1_pm.C_6_t20262	89.44

AFUA_3G12690	glfA	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WYD9]	jgi-Aspnov1-446369-fgenesh1_kg.13_#_42_#_Locus6047v1rpkm11.26	74.92
AFUA_4G06820	ecm33	Protein ecm33 [Source:UniProtKB/Swiss-Prot;Acc:Q4WNS8]	jgi-Aspnov1-442983-fgenesh1_kg.5_#_827_#_Locus141v1rpkm963.23	87
AFUA_4G11800	Alp1	Alkaline protease 1 [Source:UniProtKB/Swiss-Prot;Acc:P28296]	jgi-Aspnov1-431992-fgenesh1_pg.5_#_295	94.29
AFUA_4G12470	cpcA	BZIP transcription factor CpcA [Source:UniProtKB/TrEMBL;Acc:E9QUZ5]	jgi-Aspnov1-461084-fgenesh1_pm.5_#_230	90.87
Afua_4g14770	Helvolic acid cluster	Protostadienol synthase A [Source:UniProtKB/Swiss-Prot;Acc:Q4WR16]	jgi-Aspnov1-501426-estExt_fgenesh1_pg.C_2_t10392	42.86
Afua_4g14780	Helvolic acid cluster	Cytochrome P450 monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR17]	jgi-Aspnov1-458437-fgenesh1_pm.2_#_1359	47.7
Afua_4g14790	Helvolic acid cluster	Cytochrome P450 monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR18]	jgi-Aspnov1-458437-fgenesh1_pm.2_#_1359	92.5
Afua_4g14800	Helvolic acid cluster	Short chain dehydrogenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR19]	jgi-Aspnov1-439546-fgenesh1_kg.2_#_1427_#_Locus1689v1rpkm76.29	94
Afua_4g14810	Helvolic acid cluster	Cytochrome P450 monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR20]	jgi-Aspnov1-458435-fgenesh1_pm.2_#_1357	85.6
Afua_4g14820	Helvolic acid cluster	Transferase family protein [Source:UniProtKB/TrEMBL;Acc:Q4WR21]	jgi-Aspnov1-439545-fgenesh1_kg.2_#_1426_#_Locus3005v1rpkm38.16	91
Afua_4g14830	Helvolic acid cluster	Cytochrome P450 monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR22]	jgi-Aspnov1-458437-fgenesh1_pm.2_#_1359	47.59
Afua_4g14840	Helvolic acid cluster	Transferase family protein [Source:UniProtKB/TrEMBL;Acc:Q4WR23]	jgi-Aspnov1-439543-fgenesh1_kg.2_#_1424_#_Locus3740v1rpkm28.06	90
Afua_4g14850	Helvolic acid cluster	Extracellular 3-ketosteroid 1-dehydrogenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WR24]	jgi-Aspnov1-118200-CE118199_1683	91.79
AFUA_5G04170	hsp90	Heat shock protein 90 [Source:UniProtKB/Swiss-Prot;Acc:P40292]	jgi-Aspnov1-509883-estExt_fgenesh1_pm.C_2_t10392	99.27
AFUA_5G08570	pkaC2	cAMP-dependent protein kinase catalytic subunit, putative [Source:UniProtKB/TrEMBL;Acc:E9QXD5]	jgi-Aspnov1-100079-CE100078_981	95.71
AFUA_5G09240	cu/zn sod	Superoxide dismutase [Source:UniProtKB/Swiss-Prot;]	jgi-Aspnov1-501817-estExt_fgenesh1_pg.C_2_t20321	97.33
AFUA_5G09360	calA	Serine/threonine-protein phosphatase 2B catalytic subunit [Source:UniProtKB/Swiss-Prot;Acc:Q4WUR1]	jgi-Aspnov1-457892-fgenesh1_pm.2_#_814	97.59
AFUA_5G09580	rodA	Hydrophobin [Source:UniProtKB/Swiss-Prot;Acc:P41746]	jgi-Aspnov1-402332-estExt_Genewise1.C_2_t40215	96.43
AFUA_5G10760	mnt1	Alpha-1,2-mannosyltransferase (Kre2), putative [Source:UniProtKB/TrEMBL;Acc:Q4WV44]	jgi-Aspnov1-108431-CE108430_6570	94
AFUA_5G11230	rasA	RAS small monomeric GTPase RasA [Source:UniProtKB/TrEMBL;Acc:E9QX28]	jgi-Aspnov1-109419-CE109418_6451	99.53
AFUA_5G13300	pep1	Aspartic protease pep1 [Source:UniProtKB/Swiss-Prot;Acc:P41748]	jgi-Aspnov1-363785-e_gw1.2.2138.1	95.19
AFUA_6G04820	pabaA	Para-aminobenzoate synthase PabaA [Source:UniProtKB/TrEMBL;Acc:Q4WDI0]	jgi-Aspnov1-450180-estExt_Genemark1.C_3_t20429	88.47
AFUA_6G09570	Gliotoxin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WMK6]	jgi-Aspnov1-508952-estExt_fgenesh1_pm.C_1_t20245	30.19
AFUA_6G09580	Gliotoxin cluster	C6 finger domain protein, putative [Source:UniProtKB/TrEMBL;Acc:Q4WMK5]	jgi-Aspnov1-511168-estExt_fgenesh1_pm.C_3_t20002	82.02
AFUA_6G09590	Gliotoxin cluster	Zinc alcohol dehydrogenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WMK4]	jgi-Aspnov1-511167-estExt_fgenesh1_pm.C_3_t20001	84.89
AFUA_6G09600	Gliotoxin cluster	Zinc metallopeptidase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WMK3]	jgi-Aspnov1-502750-estExt_fgenesh1_pg.C_3_t20015	90.57

AFUA_6G09610	Gliotoxin cluster	Nonribosomal peptide syntethase 9 [Source:UniProtKB/Swiss-Prot;Acc:Q4WMK2]	jgi-Aspnov1-404104-estExt_Genewise1.C_3_t30145	77.01
AFUA_6G09620	Gliotoxin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WMK1]	jgi-Aspnov1-386801-estExt_Genewise1Plus.C_3_t30133	84.72
AFUA_6G09630	Gliotoxin cluster	C6 finger domain protein GliZ [Source:UniProtKB/TrEMBL;Acc:Q4WMK0]	jgi-Aspnov1-366288-e_gw1.3.282.1	83.75
AFUA_6G09640	Gliotoxin cluster	Aminotransferase GliI [Source:UniProtKB/TrEMBL;Acc:Q4WMJ9]	jgi-Aspnov1-458936-fgenesh1_pm.3_#_494	87.07
AFUA_6G09650	Gliotoxin cluster	Membrane dipeptidase GliJ [Source:UniProtKB/TrEMBL;Acc:Q4WMJ8]	jgi-Aspnov1-367031-e_gw1.3.965.1	91.24
AFUA_6G09660	Gliotoxin cluster	Nonribosomal peptide synthetase 10 [Source:UniProtKB/Swiss-Prot;Acc:Q4WMJ7]	jgi-Aspnov1-511160-estExt_fgenesh1_pm.C_3_t10488	92.06
AFUA_6G09670	Gliotoxin cluster	Cytochrome P450 oxidoreductase GliC [Source:UniProtKB/TrEMBL;Acc:E9RCR4]	jgi-Aspnov1-481241-MIX15903_10_44	92.59
AFUA_6G09680	Gliotoxin cluster	O-methyltransferase GliM [Source:UniProtKB/TrEMBL;Acc:Q4WMJ5]	jgi-Aspnov1-368520-e_gw1.3.829.1	93.5
AFUA_6G09690	Gliotoxin cluster	Glutathione S-transferase GliG [Source:UniProtKB/TrEMBL;Acc:A4GYZ0]	jgi-Aspnov1-458931-fgenesh1_pm.3_#_489	96.25
AFUA_6G09700	Gliotoxin cluster	Gliotoxin biosynthesis protein GliK [Source:UniProtKB/TrEMBL;Acc:E9R9Y3]	jgi-Aspnov1-429762-fgenesh1_pg.3_#_503	90.33
AFUA_6G09710	Gliotoxin cluster	MFS gliotoxin efflux transporter GliA [Source:UniProtKB/TrEMBL;Acc:E9R876]	jgi-Aspnov1-367489-e_gw1.3.2721.1	94.1
AFUA_6G09720	Gliotoxin cluster	Methyltransferase GliN [Source:UniProtKB/TrEMBL;Acc:Q4WMJ1]	jgi-Aspnov1-502743-estExt_fgenesh1_pg.C_3_t20001	84.04
AFUA_6G09730	Gliotoxin cluster	Cytochrome P450 oxidoreductase GliF [Source:UniProtKB/TrEMBL;Acc:Q4WMJ0]	jgi-Aspnov1-458927-fgenesh1_pm.3_#_485	95.44
AFUA_6G09740	Gliotoxin cluster	Thioredoxin reductase GliT [Source:UniProtKB/TrEMBL;Acc:E9RAH5]	jgi-Aspnov1-511158-estExt_fgenesh1_pm.C_3_t10480	91.92
AFUA_6G09745	Gliotoxin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WMI8]	jgi-Aspnov1-386790-estExt_Genewise1Plus.C_3_t30100	89.84
AFUA_6G10240	fos-1 (tcsA)	Sensor histidine kinase/response regulator Fos-1/TcsA [Source:UniProtKB/TrEMBL;Acc:Q4WMD9]	jgi-Aspnov1-429720-fgenesh1_pg.3_#_461	93.36
AFUA_6G11390	gel2	1,3-beta-glucanosyltransferase gel2 [Source:UniProtKB/Swiss-Prot;Acc:P0C954]	jgi-Aspnov1-398517-estExt_Genewise1.C_1_t10227	48.38
AFUA_7G04800	gprC	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WGE9]	jgi-Aspnov1-462935-fgenesh1_pm.7_#_220	94.32
AFUA_8G00170	Fumitremorgin cluster	Nonribosomal peptide synthetase 13 [Source:UniProtKB/Swiss-Prot;Acc:Q4WAW3]	jgi-Aspnov1-507323-estExt_fgenesh1_pg.C_90140	34.91
AFUA_8G00190	Fumitremorgin cluster	Cytochrome P450, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAW5]	jgi-Aspnov1-510751-estExt_fgenesh1_pm.C_3_t10008	64.49
AFUA_8G00200	Fumitremorgin cluster	O-methyltransferase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAW6]	jgi-Aspnov1-449391-estExt_Genemark1.C_3_t10007	71.39
AFUA_8G00210	Fumitremorgin cluster	Dimethylallyl tryptophan synthase FtmPT1 [Source:UniProtKB/TrEMBL;Acc:Q4WAW7]	jgi-Aspnov1-396938-estExt_Genewise1Plus.C_10_t10291	35.94
AFUA_8G00220	Fumitremorgin cluster	Cytochrome P450, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAW8]	jgi-Aspnov1-479619-MIX14281_2_12	44.02
AFUA_8G00230	Fumitremorgin cluster	Phytanoyl-CoA dioxygenase family protein [Source:UniProtKB/TrEMBL;Acc:Q4WAW9]	jgi-Aspnov1-404130-estExt_Genewise1.C_3_t30177	33.3
AFUA_8G00240	Fumitremorgin cluster	Cytochrome P450 monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAX0]	jgi-Aspnov1-419348-gm1.3695_g	42.5
AFUA_8G00250	Fumitremorgin cluster	Dimethylallyl tryptophan synthase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAX1]	jgi-Aspnov1-463437-fgenesh1_pm.8_#_19	57.76
AFUA_8G00260	Fumitremorgin cluster	F-box domain and ankyrin repeat protein [Source:UniProtKB/TrEMBL;Acc:Q4WAX2]	jgi-Aspnov1-516072-estExt_fgenesh1_pm.C_100234	27.92
AFUA_8G00370	Fumagillin cluster	Polyketide synthase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAY3]	jgi-Aspnov1-424178-gm1.8525_g	84.34

AFUA_8G00380	Fumagillin cluster	DltD N-terminal domain protein [Source:UniProtKB/TrEMBL;Acc:Q4WAY4]	jgi-Aspnov1-412667-estExt_Genewise1.C_8_t10092	96.92
AFUA_8G00390	Fumagillin cluster	O-methyltransferase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAY5]	jgi-Aspnov1-463456-fgenesh1_pm.8_#_38	96.77
AFUA_8G00400	Fumagillin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WAY6]	jgi-Aspnov1-463456-fgenesh1_pm.8_#_38	85.06
AFUA_8G00410	Fumagillin cluster	Methionine aminopeptidase 2-1 [Source:UniProtKB/Swiss-Prot;Acc:Q4WAY7]	jgi-Aspnov1-463457-fgenesh1_pm.8_#_39	91.48
AFUA_8G00420	Fumagillin cluster	C6 finger transcription factor, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAY8]	jgi-Aspnov1-434405-fgenesh1_pg.8_#_51	86.95
AFUA_8G00430	Fumagillin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WAY9]	jgi-Aspnov1-412676-estExt_Genewise1.C_8_t10101	95.54
AFUA_8G00440	Fumagillin cluster	Steroid monooxygenase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAZ0]	jgi-Aspnov1-434407-fgenesh1_pg.8_#_53	87.4
AFUA_8G00460	Fumagillin cluster	Methionine aminopeptidase [Source:UniProtKB/TrEMBL;Acc:Q4WAZ1]	jgi-Aspnov1-463462-fgenesh1_pm.8_#_44	96.22
AFUA_8G00470	Fumagillin cluster	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:Q4WAZ2]	jgi-Aspnov1-395381-estExt_Genewise1Plus.C_8_t10105	85.82
AFUA_8G00480	Fumagillin cluster	Phytanoyl-CoA dioxygenase family protein [Source:UniProtKB/TrEMBL;Acc:Q4WAZ3]	jgi-Aspnov1-445026-fgenesh1_kg.8_#_46_#_Locus3448v1rpkm31.50	94.01
AFUA_8G00490	Fumagillin cluster	PKS-like enzyme, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAZ4]	jgi-Aspnov1-445027-fgenesh1_kg.8_#_47_#_Locus11024v1rpkm1.79	72.1
AFUA_8G00500	Fumagillin cluster	Acetate-CoA ligase, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAZ5]	jgi-Aspnov1-412687-estExt_Genewise1.C_8_t10112	92.09
AFUA_8G00510	Fumagillin cluster	Cytochrome P450 oxidoreductase OrdA-like, putative [Source:UniProtKB/TrEMBL;Acc:Q4WAZ6]	jgi-Aspnov1-412689-estExt_Genewise1.C_8_t10114	94.03
AFUA_8G00520	Fumagillin cluster	Integral membrane protein [Source:UniProtKB/TrEMBL;Acc:Q4WAZ7]	jgi-Aspnov1-377876-e_gw1.8.1481.1	91.23
AFUA_8G02750	cgrA	rRNA-processing protein cgrA [Source:UniProtKB/Swiss-Prot;Acc:Q9HEQ8]	jgi-Aspnov1-454324-estExt_Genemark1.C_8_t10302	95

Table S4 Table of the terrequinone proteins from *A. nidulans* and the BLASTP hits found in *A. steynii*.

A. Nidulans terrequinone proteins	Hit protein ID in *A. steynii*	% Identity	% Coverage
TdiA - ABU51602.1	365047	71.7	101
tdiB - ABU51603.1	415228	56.2	92
tdiC - ABU51604.1	429405	52.7	88.6
tdiD - ABU51605.1	479193	69.5	96.1
tdiE - ABU51606.1	365428	56.1	97.7

Table S5

Part 1 Overview of chlorinating enzymes identified from literature. The sequence of each of these proteins were used to search for similar proteins in *A. campestris, A. candidus* and *A. taichungensis* using BLASTP comparison, but no hits were found.

Protein	Description	Reference	GenBank	*A. campestris*	*A. candidus*	*A. taichungensis*
CmaB	a chlorinating non-haem iron enzyme from Pseudomonas syringae	[1]	AAC46036.1	No hits	No hits	No hits
PrnA	a flavin dependent tryptophan halogenase from *Pseudomonas fluorescens*	[2]	AAB97504.1	No hits	No hits	No hits
PtaM	a flavin dependent halogenase from *Pestalotiopsis fici*	[3]	AGO59046.1	No hits	No hits	No hits
Thr3	an iron non heme alpha ketoglutarate dependent halogenase from Streptomyces sp. OH-5093	[4]	CCF23457.1	No hits	No hits	No hits

Part 2 Relevant InterPro domains identified by comparing the proteins listed in Suppl. Table 6 to the InterPro database using InterProScan 5 [5] and an additional word search of the database [6]. The identified InterPro domains were searched for in the annotated *A. campestris, A. candidus* and *A. taichungensis* genomes including the number of hits.

IPR ID	Description	Hits *A. campestris*	Hits *A. candidus*	Hits *A. taichungensis*
IPR000028	Chloroperoxidase Heme dependent	4	4	4
IPR001568	Ribonuclease T2	3	2	2
IPR008775	Phytanoyl-CoA dioxygenase	5	5	4
IPR010092	chlorinating enzyme FE(II)nonheme halogenase	0	0	0
IPR006905	tryptophan halogenase	0	0	0
IPR002747	SAM depdent chlorinase/fluorinase	0	0	0
IPR016119	bromoperoxidase/chloroperoxidase C-terminal	0	0	0

Part 3 Overview of the potential chlorinating proteins in the chlorflavonin candidate cluster in *A. campestris* including the best BLASTP hit in NCBI nr database and the identified InterPro IDs found using InterPro Scan 5 [5] on the protein sequences.

Protein ID	BLAST hit	InterPro Scan hit
286063	Hypothetical protein [Solirubrobacterales bacterium URHD0059] 93% coverage and 36% identity	IPR029039 – flavoprotein-like IPR005025 – NADPH-dependent FMN reductase-like
277538	Related to scytalone dehydratase [Fusarium fujikuroi IMI 58289], 100% coverage and 53% identity	IPR004235 – Scytalone dehydratase

331187	Peptidase S15/CocE/NonD, C-terminal [Penicillium expansum] 97% coverage and 41% Identity	IPR011008 Dimeric alpha-beta barrel
3988	Hypothetical protein HIM_08269 [Hirsutella minnesotensis 3608] 98% coverage and 53% Identity 3-hydroxybenzoate 6-hydroxylase 1 [Tolypocladium ophioglossoides CBS 100239] 98% coverage and 53% Identity	IPR006076 –FAD dependent oxidoreductase IPR003042 – Aromatic-ring hydroxylase IPR023753 – FAD/NAD(P) binding domain IPR002938 – FAD-binding domain

1. Vaillancourt FH, Yeh E, Vosburg D a, O'Connor SE, Walsh CT. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosyn. Nature. 2005;436:1191–4. http://www.nature.com/nature/journal/v436/n7054/pdf/nature03797.pdf. Accessed 3 Nov 2015.
2. Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ, et al. Functions Encoded by Pyrrolnitrin Biosynthetic Genes from Pseudomonas fluorescens. J Bacteriol. 1998;180:1939–43. http://jb.asm.org.proxy.findit.dtu.dk/content/180/7/1939. Accessed 3 Nov 2015.
3. Xu X, Liu L, Zhang F, Wang W, Li J, Guo L, et al. Identification of the first diphenyl ether gene cluster for pestheic acid biosynthesis in plant endophyte Pestalotiopsis fici. Chembiochem. 2014;15:284–92. doi:10.1002/cbic.201300626.
4. Fullone MR, Paiardini A, Miele R, Marsango S, Gross DC, Omura S, et al. Insight into the structure-function relationship of the nonheme iron halogenases involved in the biosynthesis of 4-chlorothreonine --Thr3 from Streptomyces sp. OH-5093 and SyrB2 from Pseudomonas syringae pv. syringae B301DR. FEBS J. 2012;279:4269–82. doi:10.1111/febs.12017.
5. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. doi:10.1093/bioinformatics/btu031.
6. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2014;43:D213-221. doi:10.1093/nar/gku1243.