

# SCIENTIFIC REPORTS



OPEN

## *Penicillium arizonense*, a new, genome sequenced fungal species, reveals a high chemical diversity in secreted metabolites

Received: 07 June 2016  
Accepted: 26 September 2016  
Published: 14 October 2016

Sietske Grijseels<sup>1,\*</sup>, Jens Christian Nielsen<sup>2,\*</sup>, Milica Randelovic<sup>1</sup>, Jens Nielsen<sup>2,3</sup>, Kristian Fog Nielsen<sup>1</sup>, Mhairi Workman<sup>1</sup> & Jens Christian Frisvad<sup>1</sup>

A new soil-borne species belonging to the *Penicillium* section *Canescentia* is described, *Penicillium arizonense* sp. nov. (type strain CBS 141311<sup>T</sup> = IBT 12289<sup>T</sup>). The genome was sequenced and assembled into 33.7 Mb containing 12,502 predicted genes. A phylogenetic assessment based on marker genes confirmed the grouping of *P. arizonense* within section *Canescentia*. Compared to related species, *P. arizonense* proved to encode a high number of proteins involved in carbohydrate metabolism, in particular hemicellulases. Mining the genome for genes involved in secondary metabolite biosynthesis resulted in the identification of 62 putative biosynthetic gene clusters. Extracts of *P. arizonense* were analysed for secondary metabolites and austerolides, pyripyropenes, tryptoquivalines, fumagillin, pseurotin A, curvulinic acid and xanthoepocin were detected. A comparative analysis against known pathways enabled the proposal of biosynthetic gene clusters in *P. arizonense* responsible for the synthesis of all detected compounds except curvulinic acid. The capacity to produce biomass degrading enzymes and the identification of a high chemical diversity in secreted bioactive secondary metabolites, offers a broad range of potential industrial applications for the new species *P. arizonense*. The description and availability of the genome sequence of *P. arizonense*, further provides the basis for biotechnological exploitation of this species.

Penicillia are important cell factories for the production of antibiotics and enzymes, and several species of the genus also play a central role in the production of fermented food products such as cheese and meat. An important characteristic of the Penicillia is their ability to produce a large number of structurally and chemically diverse secondary metabolites. These compounds include important pharmaceuticals such as the antibiotic penicillin, the cholesterol-lowering compactin and the antifungal griseofulvin, and the large diversity of bioactive compounds is a valuable reservoir for identification of new pharmaceuticals. In addition, secondary metabolites from Penicillia include mycotoxins, which are frequently found in contaminated food and feed products and pose a health risk to humans and animals. The genus *Penicillium* is large, with more than 354 species that are currently accepted<sup>1</sup>. This fungal diversity is important due to continued demand for biological sources of new enzymes and secondary metabolites<sup>2</sup> and can lead to the discovery of novel and efficient cell factories for their production.

*Penicillium arizonense* is described here as a new species belonging to *Penicillium* section *Canescentia*. Section *Canescentia* was officially described by Houbraken and Samson<sup>3</sup> and divides into two clades, one clade containing the well-known species *P. canescens* and *P. janczewskii*, and a second clade containing, *P. atrovenetum* and *P. antarcticum*, amongst others<sup>4</sup>. Members of section *Canescentia* are soil-borne and are characterized by the formation of divaricate biverticillate structures with infrequent additional branches. Phialides are simple and short (7–9 μm) with a broadly cylindrical to slightly or more definitely swollen base and a short, occasionally a more pronounced narrowed neck<sup>3</sup>.

<sup>1</sup>Department of Systems Biology, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark. <sup>2</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, SE41296 Gothenburg, Sweden. <sup>3</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.C.F. (email: [jcf@bio.dtu.dk](mailto:jcf@bio.dtu.dk))

Members of section *Canescentia* are predominantly found in forest litter and soil<sup>5</sup> and hence possess the natural ability to degrade complex substrates through secretion of a large number of diverse enzymes. Examples of degradative enzyme producers include *P. canescens*, which has been reported to efficiently produce xylanases<sup>6,7</sup> and  $\beta$ -galactosidase<sup>8</sup>, and *P. janczewskii*, which produces high amounts of xylanase,  $\beta$ -xylosidase and  $\alpha$ -L-arabinofuranosidase<sup>9</sup>. This natural production of a variety of biomass-degrading enzymes has great potential to be exploited industrially.

The most important secondary metabolite produced by members of section *Canescentia* is the industrial antifungal compound griseofulvin. Apart from *P. griseofulvum*, *P. janczewskii* and the closely related *P. nigricans* were the two subsequent species where griseofulvin was identified<sup>10,11</sup>. Later, *P. jensenii* and *P. canescens*<sup>12,13</sup> have also been identified as griseofulvin producers. Recently, griseofulvin has attracted renewed attention due to reports of complementary bioactivity in mammalian systems, including antiviral and anticancer effects<sup>14,15</sup>. Other secondary metabolites identified in *P. janczewskii* include fumagillin<sup>16</sup> amauromine (nigrifortine)<sup>17</sup>, L-Phe-L-Phe diketopiperazine<sup>18</sup>, MT 81<sup>19</sup>, the antibiotic penicillic acid<sup>12</sup>, cycloaspeptide, pseurotin A<sup>20</sup> and the indole-diterpenoids pennigritrem and penitrem A<sup>21,22</sup>, of which the latter was shown to have tumor suppressant activity in mammary cancer cells<sup>23</sup>. In 1997, patulin and roquefortine C, as well as several analogues and precursors of the latter were reported from *P. janczewskii* isolates<sup>24</sup>, however none of these have been confirmed in *P. janczewskii* or section *Canescentia* since, which also goes for the report of compactins<sup>25</sup>. Curvulinic acid has been identified in both *P. janczewskii* and *P. canescens*, while the antifungal polyketide Sch 642305 has only been detected in the latter<sup>26,27</sup>. Other secondary metabolites reported in *P. canescens* are pseurotin A, the tetrapeptide D-Phe-L-Val-D-Val-L-Tyr and aurantiamine<sup>28</sup>, however the chemical profile of this strain indicates that it could have been *P. aurantiogriseum* rather than *P. canescens*<sup>1,29</sup>. From the series, the ex-type culture of *P. canescens* was reported to produce the oxalalins and decaturin<sup>30</sup>, while the antimicrobial canescin was reported from in 1953 from *P. canescens*<sup>31</sup>. The isolate of *P. janczewskii* reported to produce penigequinolones and gliovictin<sup>32,33</sup> may have been *P. scabrosum*, as these compounds are consistently produced by the latter species<sup>34</sup>.

To fully exploit the biosynthetic capacity of filamentous fungi for the development of novel cell factories, the availability of full genome sequences is of great importance. This allows for the determination of potential biosynthetic capabilities including the identification of secondary metabolite biosynthetic gene clusters. Additionally, the genome sequences enable a knowledge based approach for the optimization of cell factories e.g. through metabolic network modelling, for the production of enzymes or secondary metabolites. Metabolic engineering strategies have previously been used with great success to redesign metabolic fluxes for secondary metabolite production in filamentous fungi<sup>35–37</sup> and for heterologous expression<sup>38–40</sup>. Enabling the transfer of genes or gene clusters to well-known expression hosts serves as an important step in the development of efficient cell factories that can be used for economically viable production of secondary metabolites in the food and pharmaceutical industry<sup>41</sup>. Heterologous expression of secondary metabolites has been amongst others successfully demonstrated in the yeasts *Saccharomyces cerevisiae*<sup>42</sup>, in the filamentous fungi *P. chrysogenum*<sup>43</sup>, *Aspergillus nidulans*<sup>44</sup>, and *A. oryzae*<sup>45,46</sup>.

Here we describe the isolation and genome sequencing of *P. arizonense*. The genome sequence is the first publicly available genome within section *Canescentia*, and hence serves as the first genomic insight into this interesting section. A functional analysis gave insights into the capacity for this species to produce degradative enzymes, which could have biotechnological implications and highlight the potential applications for *P. arizonense* as a source of industrial enzymes. We further identified putative secondary metabolite gene clusters and by coupling this with measurement of secondary metabolites we obtained valuable information on the capacity for secondary metabolite production by this species.

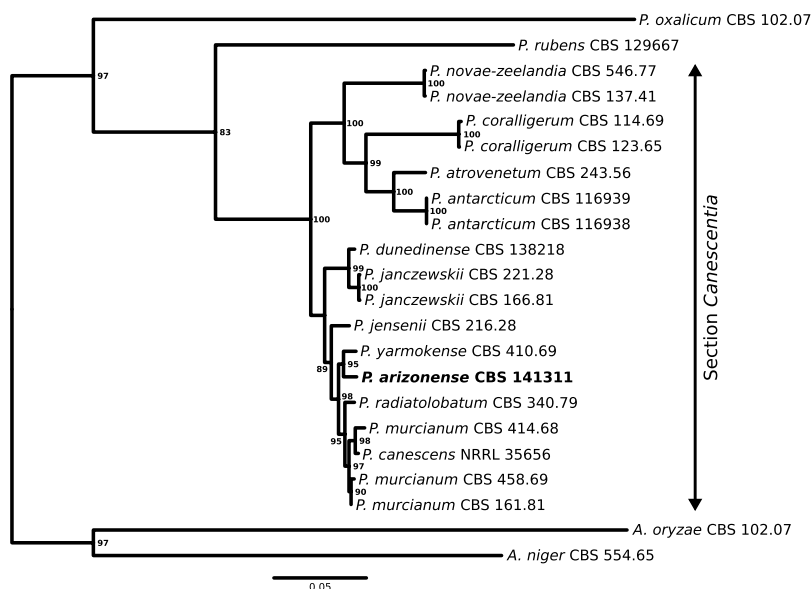
## Results and Discussion

**Genomic features.** The genome of *P. arizonense* CBS 141311<sup>T</sup> = IBT 12289<sup>T</sup> was shotgun sequenced using illumina 2500 technology (125 bp PE reads) to an approximate coverage of 154 fold. The short reads were assembled *de novo* into 396 contigs longer than 200 bp, 43 supercontigs longer than 100 kb, and the longest contig being 2.7 Mb. This resulted in a cumulative assembly length of 33.7 Mb and a contig N50 of 1 Mb. A total of 12,502 putative genes were identified in the genome including 173 tRNAs. Protein coding genes were functionally annotated by blasting CDS regions at the amino acid level against Uniprot, and 8269 proteins could be mapped to a known homolog. Quality control of the gene prediction was performed by Benchmarking Universal Single-Copy Orthologs (BUSCO)<sup>47</sup>, which assess the completeness of genomes by detecting the presence of 1438 ubiquitous single copy eukaryotic genes. All 1438 genes were found, with only 6 of them being in fragmented versions. The predicted genes covered 59.8% of the assembled genome and the majority of the genes contained at least one intron (84%) with the average number of introns being 2.2 per gene. The GC content of the assembled genome was 49.1%, which is similar to that of related species (Supplementary Table S1). Mitochondrial DNA was identified as a 28,347 bp linear DNA fragment, with a GC content of 25%, showing high similarity to the published mitochondrial genome of the somewhat related species *P. solitum*<sup>48</sup>. A total of 43 mitochondrial genes were identified including 23 tRNA genes and two ORFs orthologous to uncharacterized ORFs in the mitochondrial genome of *P. solitum*<sup>48</sup>. General genomic features of *P. arizonense* are summarized in Table 1.

**Phylogenetic analysis.** Partial nucleotide sequences of RPB2, calmodulin (*camA*),  $\beta$ -tubulin (*benA*) and the ITS sequence, from section *Canescentia* and related species were aligned, trimmed and concatenated to a final sequence of 1961 nucleotide positions, to infer a maximum likelihood phylogenetic tree (Fig. 1). The topology of the phylogram was in agreement with previous observations, grouping section *Canescentia* into two major clades<sup>4</sup>. *P. arizonense* grouped within the clade containing the well described species *P. canescens* and *P. janczewskii*, and the closest relative among the tested species was *P. yarmokense*, which represented a sister species with 95% bootstrap support.

Nuclear genome	
Size of assembled genome (Mb)	33.7
GC content (%)	49.1
Number of genes	12,502
Number of genes with introns	10,083
Mean exon per gene	3
Mean intron per gene	2
Longest gene (bp)	18,858
Mean gene length (bp)	1,614
Mean exon length (bp)	448
Mean intron length (bp)	95
Assembled genome covered by genes (%)	59.8
Number of tRNA genes	182
Mitochondrial genome	
Size of mitochondrial genome (bp)	28,347
GC content (%)	25
Number of genes	43
Number of tRNA genes	24

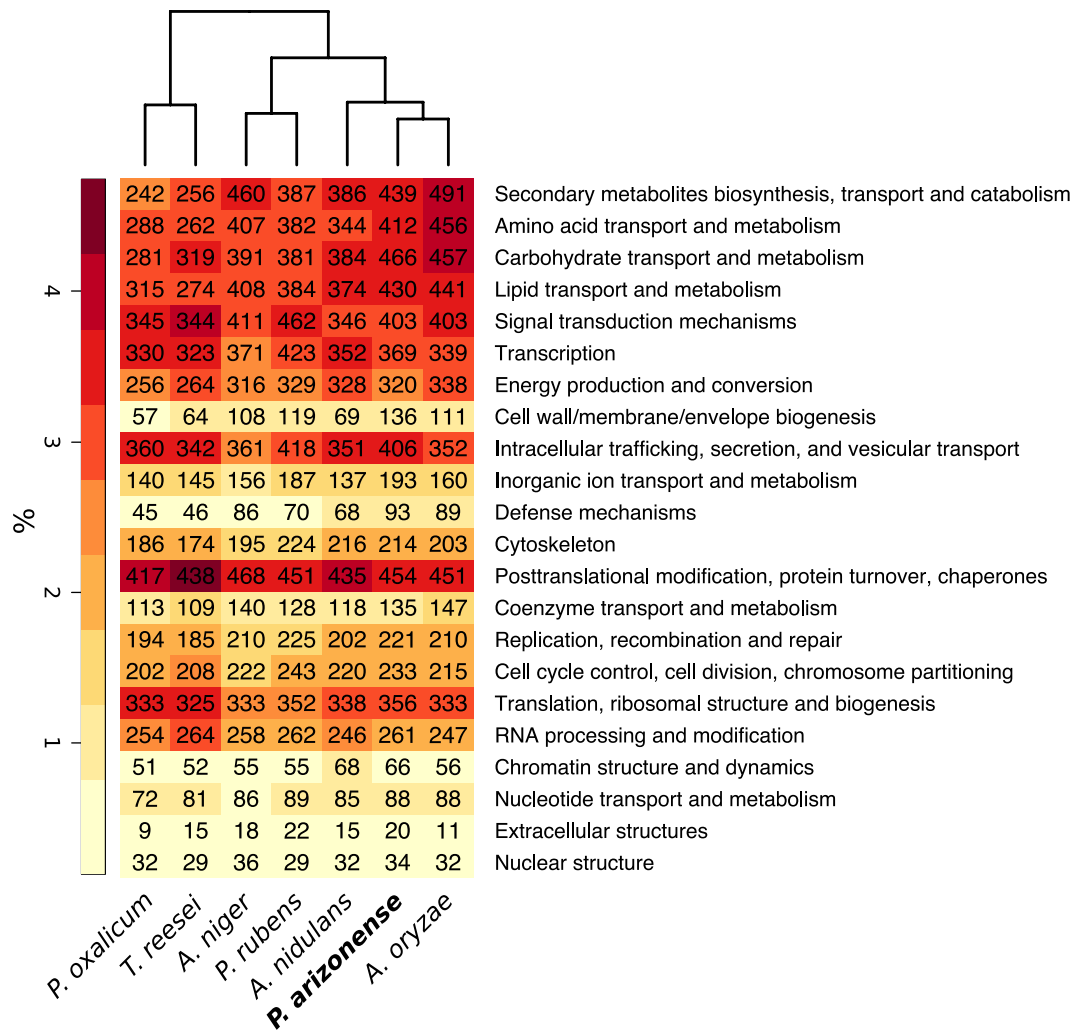
**Table 1. General features of the *P. arizonense* genome assembly.**



**Figure 1. Maximum likelihood phylogram of section *Canescentia* and related species.** The phylogram is based on partial RPB2, *camA*, *benA* and ITS sequences, and bootstrap values are given as percent of 1000 maximum likelihood trees. Only bootstrap support of more than 80% is shown.

**Functional annotation.** All proteins of *P. arizonense* were compared at a functional level to 6 related, genome sequenced fungi, selected either as being model organisms and/or because of their ability to secrete carbohydrate active enzymes (*P. rubens* (often identified as *P. chrysogenum*), *P. oxalicum*, *A. niger*, *A. nidulans*, *A. oryzae*, and *Trichoderma reesei*). Using KOG<sup>49</sup> (euKaryotic Orthologous Group) classification, we assigned functions to proteins based on sequence similarity, to determine the proportions of the proteomes allocated to different cellular functions (Fig. 2). The global pattern of protein allocation in *P. arizonense* was most similar to that of *A. oryzae*, as determined by the clustering, indicating that their ecological niches might be similar. In absolute numbers, *P. arizonense* had the highest number of proteins in the three categories, (i) “carbohydrate transport and metabolism”, (ii) “cell wall/membrane/envelope biogenesis” and (iii) “inorganic ion transport”.

The high number of proteins involved in carbohydrate metabolism in *P. arizonense* correlates with the fact that members of section *Canescentia* are known to thrive in decaying plant material where there is likely a requirement to degrade diverse and complex carbon sources. *P. arizonense* contains more genes involved in carbohydrate metabolism than any of the other species we used for the comparison, including *A. oryzae* which is known to have a diverse arsenal of carbohydrate active enzymes<sup>50</sup> and is applied for the production of industrial enzymes<sup>51</sup>. This highlights the potential for using *P. arizonense* for biotechnological applications. The value of further studying



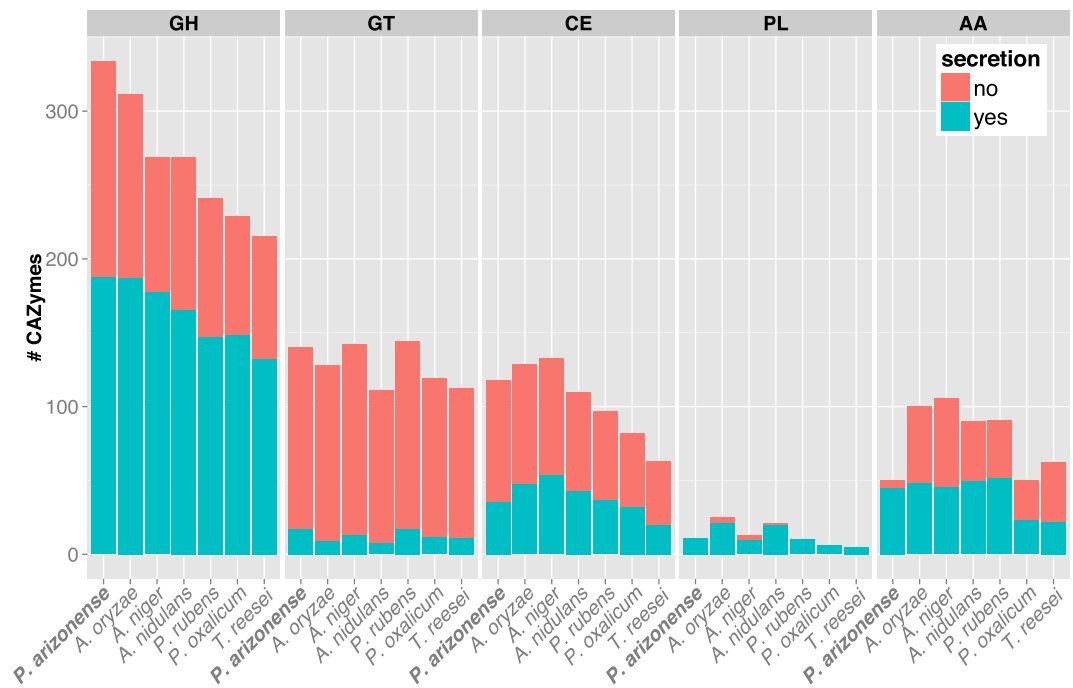
**Figure 2.** KOG (euKaryotic Orthologous Group) classification of *P. arizonense* and six related fungi. The colors represent percentage out of total number of proteins identified in the genome, and the absolute number of proteins is given as labels.

the genomics of *Penicillium* section *Canescentia* is thus emphasized, in order to elucidate whether this section as a whole is rich in carbohydrate metabolic genes or if this feature is specific to *P. arizonense*.

Furthermore, the KOG analysis showed that *P. arizonense* contains a high number of genes related to “Inorganic ion transport” which suggest that it could be an important species in the nutrient cycle in agricultural settings, where fungi are known to take up and convert inorganic phosphate to more readily accessible organic forms needed by plants. This is of major agricultural relevance since phosphate availability often is the limiting accessible nutrient in soil<sup>52</sup>. It was also noted that *P. arizonense* contains a large number of genes involved in secondary metabolism and lipid metabolism.

**Carbohydrate active enzymes.** In order to further investigate the enriched carbohydrate metabolism of *P. arizonense*, Carbohydrate Active enZymes (CAZys) were annotated and their secretion evaluated in the 7 species mentioned above (Fig. 3). In agreement with the KOG analysis showing that *P. arizonense* had a large number of carbohydrate metabolic genes, the total number of CAZymes in the species was also among the highest (668 CAZymes), only exceeded by *A. oryzae* (675 CAZymes). From the distribution of CAZy classes, the large number of CAZymes in *P. arizonense* could mainly be attributed to having a high abundance of glycoside hydrolase (GH) enzymes with 331 in total, 23 more than *A. oryzae* and 65 more than *A. niger* and *A. nidulans* (Fig. 3). The number of GH enzymes containing a secretion signal sequence was also the highest (188), although *A. oryzae* had a similar number (187).

For the carbohydrate degrading CAZy classes, polysaccharide lyases (PL), carbohydrate esterases (CE) and the glycosidic bond forming glycoside transferases (GT), *P. arizonense* had similar abundance as the other species both in total number and in number of enzymes containing secretion signals. In terms of auxiliary activity (AA) CAZymes, which are redox enzymes that work in conjunction with CAZymes aiding access to carbohydrates in plant cell walls<sup>53</sup>, *P. arizonense* contained the fewest number of enzymes together with *P. oxalicum*. However, for *P. arizonense*, the majority of these (90%) contained a secretion signal, where for the other species only 35–55% of



**Figure 3. Annotation of secreted and non-secreted CAZys in *P. arizonense* and six related fungi.** CAZymes are grouped into the classes: glycoside hydrolases (GH), glycosyl transferase (GT), carbohydrate esterases (CE), polysaccharide lyases (PL) and auxilliary activities (AA).

the AA CAZymes were annotated as secreted. How this lack of intracellular AA CAZymes affects the physiology of *P. arizonense* is unknown.

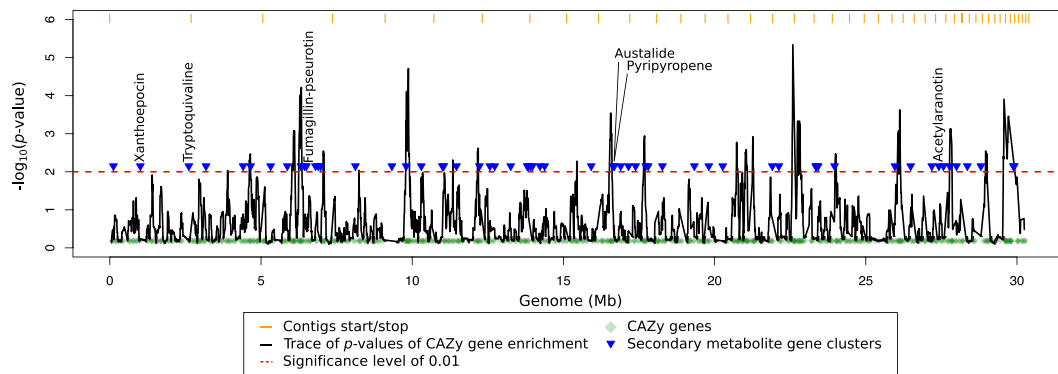
Lignocellulose degrading enzymes such as cellulases, hemicellulases and pectinases are of industrial importance for a range of applications, e.g. biomass degradation for production of biofuels, and these enzymes are mainly harboured within the GH class of CAZy. Some of the GH families were widely abundant in all species e.g. GH3 and GH18, and some were unique to a single species. *P. arizonense* contained no unique GH families, however GH29, GH39 and GH42 were seen in *P. arizonense* and each of these were found in only one other species, *A. niger*, *A. nidulans* and *P. rubens*, respectively. In the families GH1, GH65 and GH93, *P. arizonense* had a considerably higher number of enzymes than found in the compared species (see Supplementary Fig. S1). These GH families include enzymes with various catalytic activities i.e. GH1 contains mainly  $\beta$ -glucosidases and  $\beta$ -galactosidases, GH65 contains enzymes with diverse activities, but mainly phosphorylases, and GH93 contains exo- $\alpha$ -L-1,5-arabinanases. Among these enzymes, exo- $\alpha$ -L-1,5-arabinanases are of relevance for biomass conversion since they take part in the degradation of hemicellulose, and is of interest due to their potential rate limiting role in the degradation of lignocellulose<sup>54</sup>.

CAZy families specifically containing enzymes with degradative activity towards cellulose, hemicellulose and pectin were additionally quantified (see Supplementary Fig. S2). *P. arizonense* proved to contain a high number of total hemicellulases (53) together with *A. oryzae* (57), compared to the other species. The number of cellulase and pectinase encoding genes in *P. arizonense* was similar to the other species. Our results correlate with previous findings showing that closely related *P. janczewskii*, efficiently produces hemicellulose degrading enzymes, but possesses only little cellulase activity<sup>55</sup>. This feature is of interest for some processes where cellulases are unwanted, such as pulp processing, and our results indicate that it could be a property shared by *P. arizonense*.

**Secondary metabolite biosynthetic gene clusters and CAZy genes.** The genome of *P. arizonense* and 9 related *Penicillium* species were mined for putative secondary metabolite biosynthetic gene clusters using antiSMASH<sup>56</sup> (Supplementary Table S2). The total number of gene clusters in the species were highly variable ranging from 70 in *P. expansum* to 38 in *P. digitatum*, while the well-known industrially used penicillin producer, *P. rubens*, had 53. In comparison, *P. arizonense* contained a relatively high number with a total of 62 detected gene clusters (Supplementary Data S1), including 28 polyketide synthase (PKS) clusters, 16 non-ribosomal peptide synthase (NRPS) clusters, 5 terpene clusters, 3 hybrid PKS:NRPS clusters, 1 indole cluster and 1 siderophore. In particular, *P. arizonense* contained more PKS clusters than any of the compared species (Supplementary Table S2).

In *T. reesei*, CAZy genes have previously been shown to cluster into defined regions of the genome, and several of the regions enriched in CAZy genes also contained genes encoding for enzymes involved in secondary metabolism<sup>57</sup>. Since *P. arizonense* has a high abundance of both secondary metabolic gene clusters and CAZy genes, we assessed the genomic localization of these (Fig. 4), using a sliding window approach. A total of 23 genomic regions containing a significant enrichment of CAZy genes ( $p$ -value  $< 0.01$ , hypergeometric test) were found in the genome. By superimposing these regions with the predicted secondary metabolic gene clusters, it was observed that 9 CAZy enriched regions were overlapping with secondary metabolic gene clusters. Similarly,





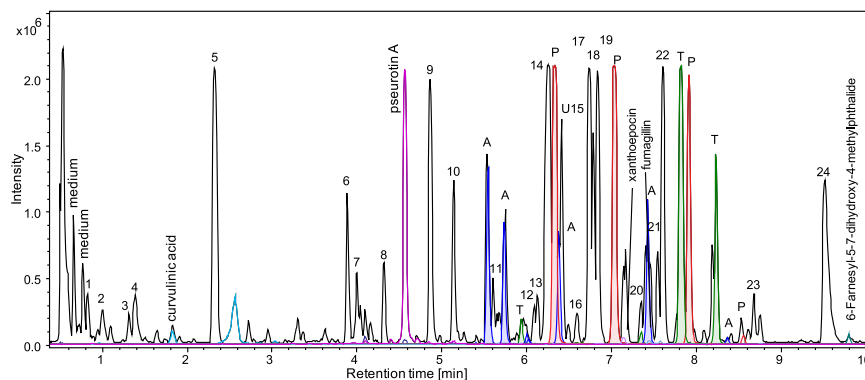
**Figure 4. Localization of CAZy genes and secondary metabolite gene clusters in the *P. arizonense* genome assembly.** CAZy gene enrichment is given as the p-value of a hypergeometric test in a sliding window of 100 kb and a step size of 10 kb. The genome assembly is visualized by concatenating supercontigs (more than 100 kb), ordered by size.

Detected compound	Most similar biosynthetic gene cluster	Orthologous genes <sup>b</sup>	Avg. similarity [% ID/% coverage]
Austalide B, J <sup>a</sup> , K, L, novel isomers C <sub>25</sub> H <sub>32</sub> O <sub>3</sub> and C <sub>26</sub> H <sub>34</sub> O <sub>3</sub>	Cluster 4: Similarity to a mycophenolic acid cluster <sup>58</sup>	3/8	59/96
6-Farnesyl-5-7-dihydroxy-4-methylphthalide <sup>a</sup>			
Pyripyropene A <sup>a</sup> , E, F, O	Cluster 5: Similarity to a pyripyropene cluster <sup>45</sup>	7/9	79/99
Tryptoquivaline C <sup>a</sup> or 27-epi-Tryptoquivaline <sup>a</sup> , G or L, I, M or 27-epi-Nortryptoquivaline	Cluster 3: Similarity to a tryptoquivaline cluster <sup>67</sup>	13/13	74/96
Fumagillin <sup>a</sup>	Cluster 34: Similarity to an intertwined fumagillin/pseurotin cluster <sup>71,95</sup>	16/16	82/95
Pseurotin A <sup>a</sup>			
Xanthoepocin	Cluster 2: Similarity to an aurofusarin cluster <sup>79,80</sup>	7/11	49/96
Curvulinic acid <sup>a</sup>	n.d.	n.a.	n.a.
n.d.	Cluster 28: Similarity to an acetylaranotin cluster <sup>83</sup>	9/9	68/91

**Table 2. Summary of secondary metabolites identified in extracts of *P. arizonense* and predicted biosynthetic gene clusters.** All compounds were found in extract of *P. arizonense* IBT 12295, 12287 and 12289. <sup>a</sup>confirmed with a standard. <sup>b</sup>Number of genes in the predicted *P. arizonense* gene cluster which are orthologous (more than 30% identity and 50% coverage) to a gene in the reference gene cluster. For detailed information about the detected gene clusters see Supplementary Data S1 and S2.

Martinez *et al.*<sup>57</sup> identified 28 enriched CAZy gene regions in *T. reesei*, where 5 of these contained either a PKS or an NRPS<sup>57</sup>. Although most secondary metabolite gene clusters do not cluster with regions enriched in CAZy genes in both *P. arizonense* and *T. reesei*, several overlaps were seen (Fig. 4). Among the co-localized secondary metabolite gene clusters we found in *P. arizonense*, one was predicted to encode the synthesis of austalides and another pyripyropenes (see section below and Table 1). It is intriguing to speculate whether these secondary metabolites could be regulated commonly with the CAZy genes in the region, and if their products have synergistic effects.

**Secondary metabolites.** To further investigate the identified diversity in secondary metabolite gene clusters at a chemical level, *P. arizonense* was, after a pre-screening (data not shown), grown on three different solid media and analyzed using liquid chromatography combined with UV/Vis spectroscopy and high resolution mass spectrometry for secondary metabolite profiling. Seven different compounds or families of compounds were identified in the crude extract of *P. arizonense* (Table 2, Supplementary Table S3, Supplementary data S4 and Supplementary Fig. S3). The major peaks identified, belonged to the families of pyripyropenes, austalides and tryptoquivalines. Except for curvulinic acid, pseurotin A and fumagillin, none of the secondary metabolites identified in the extract had previously been found in section *Canescentia*. The compounds compactin, roquefortines, mycelianamide, penigequinolones, chrysogines, and oxalicens/decaturnins could not be chemically detected in *P. arizonense*, even though they have been reported to be produced in other species in section *Canescentia*. Many of the peaks in the chromatogram could not be matched with any compounds in the database Antibase (Fig. 5) (Wiley-VCH, Weinheim, Germany), and could be interesting targets for isolation and structure elucidation by NMR. Based on homology to known gene clusters, the predicted clusters in *P. arizonense* were connected to the detected compounds when possible (Table 2 and Supplementary Data S2). All compounds except curvulinic acid could putatively be connected to a predicted secondary metabolite gene cluster in the genome.



**Figure 5.** Base peak chromatogram and extracted ion chromatograms of extract of *P. arizonense* grown on CYA medium. Black: base peak chromatogram, blue: extracted ion chromatograms of molecular ions of austrialides (A), red: extracted ion chromatograms of molecular ions of pyripyropenes (P), green: extracted ion chromatograms of molecular ions of tryptoquivalines (T). Major peaks that could not be identified: 1–24.

**Austrialides.** A total of six austrialides were identified based on accurate mass, UV/Vis, and very similar MS/HRMS spectra with  $m/z$  207.065 dominating. The F J isomer could be unambiguously identified by comparison to reference standards, while the remaining five major austrialides were only tentatively identified. The austrialide gene cluster was not previously described in the literature, but one of the predicted gene clusters in the genome of *P. arizonense* showed similarity to the mycophenolic acid gene cluster of *P. brevicompactum*<sup>58</sup>. Mycophenolic acid is a meroterpenoid consisting of an acetate-derived phthalide nucleus and a terpene-derived side chain. This phthalide structure is also present in the austrialides and it is therefore likely that mycophenolic acid and austrialides have similar biosynthetic genes. In a putative austrialide gene cluster in *P. arizonense*, orthologs were found to the genes *mpaC*, *mpaD*, and *mpaA* from the mycophenolic acid cluster in *P. brevicompactum*, corresponding to the genes responsible for the biosynthesis of the phthalide intermediate, 6-farnesyl-5,7-dihydroxy-4-methylphthalide<sup>58</sup>. This phthalide was also identified in the extract of *P. arizonense*, and confirmed by comparison of HRMS, UV/Vis and MS/HRMS to a reference standard. The austrialides are a family of related meroterpenoid metabolites that were isolated for the first time from maize cultures of *A. ustus*<sup>59</sup>. Recently, it was reported that they also were produced by *P. thomii* KMM 4645 and *P. lividum*<sup>60</sup>.

**Pyripyropenes.** Four compounds with accurate masses and UV spectra corresponding to the family of pyripyropenes were detected; pyripyropene E, F, O, and A. The latter was further confirmed by comparison of retention time and tandem MS (MS/HRMS) spectra to a reference standard. The other pyripyropenes had UV spectra and MS/HRMS fragmentation patterns similar to pyripyropene A. The biosynthetic gene cluster of pyripyropene was identified in *A. fumigatus*<sup>45</sup> and have later been described in *P. coprobium* as well<sup>29,61</sup>. One of the predicted PKS clusters in *P. arizonense* contained orthologous genes to 7 of the 9 cluster members of the *A. fumigatus* pyripyropene gene cluster. For the two missing genes the blast analysis suggested a fusion of the *A. fumigatus* *pyr1* and *pyr2* as well as *pyr4* and *pyr5* in the *P. arizonense* cluster, and this would correspond to a full conservation of the gene order in the two clusters. Pyripyropenes were first isolated from *A. fumigatus*<sup>62</sup> and later they were found to also be produced by *P. reticulisporum*<sup>63</sup>, *P. coprobium*, *P. concentricum* and *P. coprophilium*<sup>29</sup>. Pyripyropenes have attracted interest since they are highly selective toward, Acyl-CoA:cholesterol acyltransferase, a potential therapeutic target for the treatment or prevention of hypercholesterolemia and atherosclerosis<sup>64</sup>, and showed potent anti-proliferative activity against Human Umbilical Vein Endothelial Cells (HUVECs)<sup>65</sup>. It is interesting to note that while *P. arizonense* produces pyripyropenes, the closely related species *P. canescens* produces the chemically related oxalicens/decaturins<sup>30</sup>.

**Tryptoquivalines.** The presence of a tryptoquivaline with elemental composition  $C_{29}H_{30}N_4O_7$  was confirmed with retention time, accurate mass and MS/HRMS matching the reference standard. This elemental composition corresponds with tryptoquivaline C and 27-epi-tryptoquivaline, however because no stereochemistry was provided by the manufacturer of the standard, they could not be distinguished. Three other peaks had UV spectra and fragmentation patterns similar to the tryptoquivaline standard. However, there are several tryptoquivalines with the same elemental compositions and with only a limited number of tryptoquivaline reference standards; these compounds could only be identified as members of the tryptoquivaline group. The closely related compounds, tryptoquialanins, have been identified in *P. digitatum*<sup>66</sup> and the gene cluster was described in *P. lanosocoeruleum* (formerly *P. aethiopicum*)<sup>67</sup>. A predicted NRPS cluster in *P. arizonense* was highly similar to the tryptoquialanin cluster of *P. lanosocoeruleum* and contained orthologs of all 13 biosynthetic genes, as well as conservation of the gene synteny except for two genes which have swapped position. Tryptoquivalines are tremorgenic and were first identified in *Aspergillus* species<sup>68,69</sup> and later in *P. jamesonlandense*<sup>70</sup>.

**Fumagillin and pseurotin A.** The presence of fumagillin in the extract of *P. arizonense* was confirmed by comparison of HRMS, UV/Vis and MS/HRMS to a reference standard. It has been shown that the biosynthetic genes

of fumagillin are, together with the genes of the compound pseurotin A, located in an intertwined supercluster in *A. fumigatus*<sup>71</sup>. Based on a comparison of HRMS, UV/Vis and MS/HRMS to a reference standard, pseurotin A was also confirmed in the *P. arizonense* extract. A single gene cluster highly similar to the intertwined fumagillin-pseurotin gene cluster in *A. fumigatus*, was found in the genome of *P. arizonense*, encoding orthologs of all enzymes for the fumagillin and pseurotin A biosynthesis. The gene synteny was also fully conserved in the two clusters including 4 co-localized genes that have not been connected to the biosynthesis of fumagillin or pseurotins so far. Similar intertwined fumagillin-pseurotin clusters have been seen in closely related *A. fischeri* (formerly *Neosartorya fischeri*) and in the more distantly related *Metarhizium anisopliae*, but with a less conserved gene organization compared to *A. fumigatus* and *P. arizonense*. Interestingly, several other members of section *Canescentia* produce fumagillin, but only *P. janczewskii* has been reported to produce pseurotin A<sup>20</sup> as well, suggesting that it also might have the intertwined cluster. This finding suggest that an evolutionary pressure exists in some species to keep the biosynthetic genes for fumagillin and pseurotin A co-localized, possibly related to their co-regulation, which is governed by the LaeA regulated transcription factor FapR in *A. fumigatus*<sup>71</sup>. Fumagillin was first identified in *A. fumigatus*<sup>72</sup> and later it has been seen in *P. scabrosum*<sup>12</sup>, *P. jensenii*, *P. nigricans*<sup>73</sup>, *P. janczewskii*<sup>16</sup>, all belonging to section *Canescentia*, as well as an undescribed species also allocated to section *Canescentia*<sup>74</sup>. It has antibiotic and antifungal activity and it was also discovered to have anti-cancer properties<sup>75</sup>. Pseurotin A is a heterospirocyclic secondary metabolite originally isolated from *Pseudeurotium ovale*<sup>76</sup>. Pseurotin A has been found to induce cell differentiation of PC12 neuronal cells, highlighting the compound as a potential tool for studying the mechanism of neurite formation<sup>77</sup>. The production of immunoglobulin E (IgE) is suppressed by pseurotin A, suggesting its use as an interventional tool for studies of IgE-mediated systemic allergic response mechanisms<sup>78</sup>.

**Xanthoepocin.** For xanthoepocin there was no analytical standard available but previous dereplication, MS/HRMS fragmentation and the distinctive UV spectrum, previously described by Igarashi *et al.* (2000) (see Supplementary Fig. S4), showed its presence in the extract. The xanthoepocin cluster is not known, but a cluster of the related compound, aurofusarin have been shown to consist of 11 genes in *F. graminearum*<sup>79,80</sup>. One *P. arizonense* PKS cluster contained orthologs of 7 of these genes including the PKS, and hence might be responsible for the xanthoepocin biosynthesis. Xanthoepocin has antibiotic activity and was previously identified from *P. simplicissimum*<sup>81</sup> and *P. excelsum*<sup>82</sup>.

**Curvulinic acid.** The presence of curvulinic acid was confirmed by comparison of HRMS, UV/Vis and MS/HRMS to a reference standard. The gene cluster for curvulinic acid is not described in the literature and amongst the 28 putative PKS clusters found in *P. arizonense* it was not possible to assign any of them to the curvulinic acid biosynthesis. Curvulinic acid was previously shown to be produced by another member of section *Canescentia*, *P. canescens*<sup>26</sup>.

In addition to the detected compounds, a predicted NRPS cluster in *P. arizonense* showed similarity to an acetylaranotin gene cluster from *A. terreus*, with orthologs of all 9 genes<sup>83</sup>. Acetylaranotin is an epipolythiodioxopiperazine, which is a class of secondary metabolites containing di- or polysulfide bridges. None of the peaks in the extracts of *P. arizonense* showed isotopic patterns consistent with sulphur, hence, this putative acetylaranotin cluster is assumed to be silent under the conditions tested in this study.

**Taxonomy.** *Penicillium arizonense* Frisvad, Grijseels and J.C. Nielsen, sp. nov.

**Mycobank MB 817128.** Type: Herb. C-F-101845, cultures ex type IBT 12289 = CBS 141311, from a sample of dry red soil, south rim of Grand Canyon, Grand Canyon Village, Arizona, USA (36°, 3'22.31" N; 112° 7'30.73" W), Per V. Nielsen, July 1990, fungus isolated by dilution plating by J.C. Frisvad; additional cultures: IBT 12285 = CBS 141312; IBT 12287 = CBS 141313 from the same source as the culture ex type, but not of the same clone.

**Morphological description.** Macromorphology. Colony diameter on CYA agar (Czapek Yeast Autolysate agar) after one week in darkness at 25 °C: 28–35 mm, MEA (Malt Extract Agar): 11–29 mm, MEA Oxoid: 18–28 mm, YES agar (Yeast Extract Sucrose agar): 28–42 mm, OAT agar (oatmeal agar): 20–34 mm, CYA at 37 °C: no growth, CREA agar (Creatine sucrose agar): 13–19 mm, weak growth, no acid production. Colonies on CYA agar, moderate to good sporulation, reverse dark orange brown to dark brown, conidia *en masse* dull green to grey green, exudate droplets produced, large, light yellow, colonies on MEA, moderate to good sporulation, colony reverse orange to orange brown, the colour diffusing into the agar, colonies on YES moderate sporulation, reverse red brown to violet brown, colonies on oat meal agar, good sporulation, reverse yellow to orange. Pictures of colonies grown on CYA, YES and MEA are shown in Fig. 6.

**Micromorphology.** Penicilli irregularly biverticillate and strongly divaricate, borne from aerial hyphae, often appearing as a divergent tetrad of metulae, but also bearing intercalary metulae and even a lower ramus with metulae or very irregular, some metulae appearing as monoverticillate independent penicilli, stipes varying in length and character, mostly 50–400 × 2.5–3.2 μm, smooth-walled, metulae cylindrical to slightly apically swollen, 8–16 μm × 2–3 μm, phialides ampulliform in verticils of 5–8, 6.5–8 μm × 2.2–2.8 μm with a distinct neck, conidia globose, smooth to finely roughened wall ornamentation, arranged in poorly defined columns. Photomicrographs of conidiophores and conidia are shown in Fig. 6.

## Methods

**Strain isolation and preservation.** The isolates IBT 12285, IBT12287 and IBT12289 were obtained from soil in Grand Canyon, south Rim, USA, Arizona in July 1990. The cultures were contained in the culture





**Figure 6.** Micro- and macro morphology of *P. arizonense*. Colonies after one week (a) CYA (b) YES (c) MEA, (d–h) conidiophores (i) conidia.

collection of DTU, Department of Systems Biology, Denmark (IBT). The strains have also been deposited at CBS (see under description of the new species). Strain IBT 12289 was used for genome sequencing and for this study.

**Genomic DNA extraction.** *P. arizonense* was grown for 7 days at 25 °C on CYA agar to induce sporulation. Conidiophores were harvested in water with 0.1% tween (v/v) and 0.9% NaCl (w/v), and used for inoculation ( $10^9$  spores/l) of 500 ml baffled Erlenmeyer shake flasks with a working volume of 150 ml CYA medium. Cultivation was carried out for 5 days at 25 °C with orbital shaking (150 rpm). Mycelium was isolated from the remaining culture broth by filtration through mira-cloth and subsequently lyophilized. Extraction of genomic DNA for sequencing is described in Supplementary Methods S1.

**Genome sequencing and assembly.** Extracted DNA was sequenced using illumina 2500 technology, yielding 125 bp paired end reads with an average insert size of 350 bp. The raw fastq files were quality checked and assembled with various assembly tools (ABYSS<sup>84</sup> (v 1.5.2), SOAPdenovo (v. 2.04), SPAdes (v. 3.5.0) and MIRA (v. 4.0.2)). De bruijn graph based assemblers were tested with k-mer values between 55–117, and the quality of the assemblies was assessed using QUAST<sup>85</sup> and FRCalign<sup>86</sup>. Finally the assembly of ABYSS with a k-mer size of 93 was chosen as the best assembly and was used for the further analysis.

**Genome annotation.** A custom gene prediction pipeline using the Maker<sup>87</sup> package was applied to combine evidence data (protein homology, transcripts, repeats) and *ab initio* predictions into gene annotations (for details see Supplementary Methods S2). Predicted genes were functionally annotated based on the protein sequence

queried against Swiss-prot to retrieve gene name and protein functions following the best-hit principle when the BLAST e-value was inferior to  $1e^{-6}$ . In addition, tRNAscan<sup>88</sup> (v. 1.3.1) was used to predict tRNAs. To identify the mitochondrial genome, a nucleotide blast of the genome assembly against the mitochondrial genome of *P. solitum*<sup>48</sup> was conducted to unambiguously identify one highly similar contig most likely representing the mitochondrial genome of *P. arizonense*. Mitochondrial genes were annotated using the MFannot tool (<http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>).

For a functional comparison of the predicted proteome of *P. arizonense* we downloaded the proteomes of 6 related species and classified all proteins into KOGs (retrieved on 2015-08-21 from: [ftp.ncbi.nih.gov/pub/mmdb/cdd/little\\_endian](ftp.ncbi.nih.gov/pub/mmdb/cdd/little_endian)) using rpsblast, with an e-value cut-off set to 0.01 (see list of species in Supplementary Table S1). Proteins of each species were subsequently mapped to the CAZy database using family specific profile hidden markov models, downloaded from dbCAN<sup>89</sup> (v. 4.0). For each CAZyme, secretion was predicted by evaluating the presence of a secretion signal sequence using signalP<sup>90</sup> (v. 4.1) with default parameters.

Secondary metabolite biosynthetic gene clusters were predicted using antiSMASH<sup>56</sup> (v. 3.0.4) in *P. arizonense* and related *Penicillium* genomes for comparison (see list of species in Supplementary Table S2). Predicted gene clusters in *P. arizonense* were connected to known pathways based on the antiSMASH output and custom BLAST analysis. Two genes were considered orthologous if a protein BLAST resulted in a global identity of more than 30% and a coverage more than 50%.

**Phylogenetic analysis.** The phylogenetic relationship within section *Canescentia* and related species was inferred based on partial sequences of the genes RPB2 (602 bp), CaM (427 bp), BenA (410 bp), as well as the Internal Transcribed Spacer sequence (ITS) (522 bp). The sequences were extracted with accession numbers from NCBI (see Supplementary Data S3), and the corresponding loci in the *P. arizonense* genome were identified with a nucleotide BLAST using the sequences of related species as query. A nucleotide multiple sequence alignment was generated using MUSCLE<sup>91</sup> (v. 3.8.31) with default parameters and subsequently, poorly aligned regions were removed using Gblocks<sup>92</sup> (v. 0.91b) with parameters  $-b2 = 50$   $-b5 = a$ , and concatenated into a single sequence of 1961 nucleotide positions. A maximum likelihood phylogenetic tree was inferred using PhyML<sup>93</sup> (v. 20120412), with GTR + I + G as substitution model. Support in nodes was calculated with 1000 bootstrap replicates and only bootstrap support of more than 80% is shown.

**Secondary metabolite analysis.** The cultures were grown on CYA (Czapek Yeast Autolysate agar: 5 g/l yeast extract, 35 g/l Czapek dox broth, 1 ml/l trace metal solution, 15 g/l agar), YES (Yeast Extract Sucrose agar, 20 g/l yeast extract, 150 g/l sucrose, 0.5 g/l  $MgSO_4 \cdot 7H_2O$ , 1 ml/l trace metal solution, 15 g/l agar), OAT (30 g/l oatmeal, 1 ml/l trace metal solution, 15 g/l agar), for 7 days at 25 °C. Three agar plugs were sampled from one colony on each medium and 1.0 ml of extraction solvent, isopropanol:ethylacetate (1/3) containing 1% formic acid, was added. After ultra-sonication for 1 h the extract was transferred to a clean vial, evaporated to dryness and redissolved in 100  $\mu$ l methanol. After centrifugation for 5 min the supernatant was directly used for chemical analysis.

Secondary metabolite profiling was done by UHPLC-DAD-TOFMS on a maXis HD orthogonal acceleration quadrupole time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization (ESI) source and connected to an Ultimate 3000 UHPLC system (Dionex, Thermo Scientific, Dionex, Sunnyvale, CA). The column used was a Kinetex 2.6  $\mu$ m C18, Å 00D7 2.1 mm (Phenomenex, Torrance, CA) maintained at 40 °C with a flow rate of 0.4 ml/min. A linear gradient system composed of 20 mmol/L formic acid in water, and 20 mmol/L formic acid in acetonitrile was used, starting from 10% (v/v) acetonitrile and increased to 100% in 10 min, maintaining this rate for 3 min before returning to the starting conditions within 0.1 min and maintaining these for 2.4 min before the following run. TOFMS was performed in ESI<sup>+</sup> with a data acquisition range of 10 scans per second at  $m/z$  100–1,000, switching between 0 and 20 eV fragmentation energy. The TOFMS was calibrated at the start of each analytical run using Bruker Daltonics high precision calibration algorithm with the use of the internal standard sodium formate. UV/Vis spectra were collected at wavelengths from 200 to 700 nm. Identification of secondary metabolites was performed using aggressive dereplication of the full HRMS data, and pseudo MS/MS data from the 20 eV fragmentation trace, using a search list of compounds based on former taxonomic identification and a manual search of major peaks in the internal library. This library consists of 1500 compounds of which 95% are fungal secondary metabolites<sup>94</sup>. Compounds were confirmed by comparison of HRMS, UV/Vis and MS/HRMS to a reference standard.

True tandem MS/HRMS spectra were made on an Agilent 1290 UHPLC system (Agilent Technologies, Torrance, CA) using a similar separation system and coupled to an Agilent 6545 QTOF where MS/HRMS spectra were obtained at fixed collision-induced dissociation (CID) energies of 10, 20, and 40 eV<sup>94</sup> and matched to the internal library of approx. 1500 reference standards and previously tentatively identified compounds.

**Morphological analysis.** Colony descriptions were based on 7 days growth of cultures CYA agar, Blakeslee malt extract agar (MEA), Oxoid malt extract agar (MEA-Ox), YES agar, OAT agar, and creatine sucrose agar incubated in darkness at 25 °C<sup>1</sup>. Micromorphological features were examined on MEA.

## References

1. Visagie, C. M. *et al.* Identification and nomenclature of the genus *Penicillium*. *Stud. Mycol.* **78**, 343–371 (2014).
2. Chamberg, F. S. & Valencia, E. Y. Fungal biodiversity to biotechnology. *Appl. Microbiol. Biotechnol.* **100**, 2567–2577 (2016).
3. Houbraken, J. & Samson, R. A. Phylogeny of *Penicillium* and the segregation of *Trichocomaceae* into three families. *Stud. Mycol.* **70**, 1–51 (2011).
4. Visagie, C. M. *et al.* *Aspergillus*, *Penicillium* and *Talaromyces* isolated from house dust samples collected around the world. *Stud. Mycol.* **78**, 63–139 (2014).
5. Domsch, K. H., Gams, W. & Anderson, T.-H. *Compendium of soil fungi. Second edition.* (IHW-Verlag, Eching, 2007).
6. Bakri, Y., Jacques, P. & Thonart, P. Xylanase production by *Penicillium canescens* 10-10c in solid-state fermentation. *Appl. Biochem. Biotechnol.* **105–108**, 737–748 (2003).

7. Gaspar, A., Cosson, T., Roques, C. & Thonart, P. Study on the production of a xylanolytic complex from *Penicillium canescens* 10–10c. *Appl. Biochem. Biotechnol.* **67**, 45–58 (1997).
8. Zheretsov, N. A., Dorokhov, V. V., Suvorova, I. M. & Cheremnova, T. A. Study of the biosynthesis of beta galactosidase from *Penicillium-canescens*. *Biokhimiya* **1986**, 50–56 (1986).
9. Terrasan, C. R. F., Temer, B., Duarte, M. C. T. & Carmona, E. C. Production of xylanolytic enzymes by *Penicillium janczewskii*. *Bioresour. Technol.* **101**, 4139–4143 (2010).
10. Brian, P. W., Curtis, P. J. & Hemming, H. G. A substance causing abnormal developmet of fungal hyphae produced by *Penicillium janczewskii* Zal. *Trans. Br. Mycol. Soc.* **32**, 30–33 (1949).
11. MacMillan, J. Dechlorogriseofulvin - a metabolic product of *Penicillium griseofulvum* Dierckx and *Penicillium janczewskii* Zal. *Chem. Ind.* **1951**, 7–9 (1951).
12. Frisvad, J. C. & Filtenborg, O. Revision of *Penicillium* subgenus *Furcatum* based on secondary metabolites and conventional characters. In Samson, R. A. & Pitt, J. I. (eds): *Modern concepts in Penicillium and Aspergillus classification*. Plenum Press, New York. 159–170 (1990).
13. Wang, Y., Wang, G., Wang, L., Xu, X., Xia, J., Huang, X. & Wu, Y. Z. C. Isolation and identification of an endophytic fungus of *Polygonatum cyrtoneuma* and its antifungal metabolites. *Acta Microbiol. Sin.* **50**, 1036–1043 (2010).
14. Yuan-Soon, Ho, J.-S. D. *et al.* Griseofulvin Potentiates Antitumorigenesis Effects of Nocodazole Through Induction of Apoptosis and G2/M Cell. *Science (80-)*. **401**, 393–401 (2001).
15. Petersen, A. B., Rønneest, M. H., Larsen, T. O. & Clausen, M. H. The chemistry of griseofulvin. *Chem. Rev.* **114**, 12088–12107 (2014).
16. Kwon, J. Y. *et al.* cis-fumagillin, a new methionine aminopeptidase (type 2) inhibitor produced by *Penicillium* sp. F2757. *J. Antibiot. (Tokyo)*. **53**, 799–806 (2000).
17. Laws, I. & Mantle, P. G. Nigrifortine, a diketopiperazine metabolite of *Penicillium nigricans*. *Phytochemistry* **24**, 1395–1397 (1985).
18. Birkinshaw, J. H. & Hohammed, Y. S. Studies in the biochemistry of microorganisms. 111. The production of l-phenylalanine anhydride (cis-1-3,6-dibenzyl-2,5-dioxopiperazine) by *Penicillium nigricans* (Bainier) Thom. *Biochem J.* **85**, 523–527 (1962).
19. Gupta, M., Chatterjee, T., Sengupta, S. & Majumdar, S. K. Structure of a new mycotoxin MT-81. *Indian J. Chem. Sect. B Org. Chem. Incl. Med. Chem.* **23B**, 393–394 (1984).
20. Schmeda-Hirschmann, G., Hormazabal, E., Rodriguez, J. A. & Theoduloz, C. Cycloaspeptide A and pseurotin A from the endophytic fungus *Penicillium janczewskii*. *Zeitschrift für Naturforschung. C, J. Biosci.* **63c**, 383–388 (2008).
21. Mantle, P. G. & Penn, J. A Role for Paxilline in the Biosynthesis of Indole-Diterpenoid Penitrem Mycotoxins. *J. Chem. Soc. Perkin Trans. 1* **1989**, 1539–1540 (1989).
22. Penn, J., Biddle, J. R., Mantle, P. G., Bilton, J. N. & Sheppard, R. N. Pennigritrem, a naturally-occurring penitrem-A analog with novel cyclization in the diterpenoid moiety. *J. Chem. Soc. Trans. 1* **1992**, 23–26 (1992).
23. Sallam, A. A. *et al.* Indole diterpene alkaloids as novel inhibitors of the Wnt/beta-catenin pathway in breast cancer cells. *Eur. J. Med. Chem.* **70**, 594–606 (2013).
24. Kozlovskii, A. G., Vinokurova, N. G., Zhelifonova, V. P. & Adanin, V. M. Secondary metabolites of fungi belonging to the species *Penicillium janczewskii*. *Appl. Biochem. Microbiol.* **33**, 61–64 (1997).
25. Chu, Y., Yang, X. & Peng, Y. A new producer of mevastatin. *Zhongguo Kangshengsu Zazhi* **24**, 4–6 (1999).
26. Nicoletti, R., Lopez-Gresa, M. P., Manzo, E., Carella, A. & Ciavatta, M. L. Production and fungitoxic activity of Sch 642305, a secondary metabolite of *Penicillium canescens*. *Mycopathologia* **163**, 295–301 (2007).
27. Aldridge, W. B. & Turner, D. C. *Fungal metabolites*. Vol. II. (Academic Press, New York, 1983).
28. Bertinetti, B. V., Peña, N. I. & Cabrera, G. M. An antifungal tetrapeptide from the culture of *Penicillium canescens*. *Chem. Biodivers.* **6**, 1178–1184 (2009).
29. Frisvad, J. C. & Samson, R. A. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*: A guide to identification of food and air-borne terverticillate Penicillia and their mycotoxins. *Stud. Mycol.* **49**, 1–173 (2004).
30. Yaegashi, J., Romsdahl, J., Chiang, Y.-M. & Wang, C. C. C. Genome mining and molecular characterization of the biosynthetic gene cluster of a diterpenic meroterpenoid, 15-deoxyoxalidine B, in *Penicillium canescens*. *Chem. Sci.* **6**, 6537–6544 (2015).
31. Brian, P. W., Hemming, H. G., Moffatt, J. S. & Unwin, C. H. Canescin, an antibiotic produced by *Penicillium canescens*. *Trans. Br. Mycol. Soc.* **36**, 243–247 (1953).
32. Schmeda-Hirschmann, G., Hormazabal, E., Astudillo, L., Rodriguez, J. & Theoduloz, C. Secondary metabolites from endophytic fungi isolated from the Chilean gymnosperm *Prumnopitys andina* (Lleuque). *World J. Microbiol. Biotechnol.* **21**, 27–32 (2005).
33. He, J. *et al.* Diastereomeric quinolinone alkaloids from the marine-derived fungus *Penicillium janczewskii*. *J. Nat. Prod.* **68**, 1397–1399 (2005).
34. Larsen, T. O., Smedsgaard, J., Frisvad, J. C., Anthoni, U. & Christophersen, C. Consistent production of penigequinolone A and B by *Penicillium scabrosum*. *Biochem. Syst. Ecol.* **27**, 329–332 (1999).
35. Harris, D. M. *et al.* Engineering of *Penicillium chrysogenum* for fermentative production of a novel carbamoylated cephem antibiotic precursor. *Metab. Eng.* **11**, 125–137 (2009).
36. Thykaer, J., Christensen, B. & Nielsen, J. Metabolic network analysis of an adipoyl-7-ADCA-producing strain of *Penicillium chrysogenum*: elucidation of adipate degradation. *Metab. Eng.* **4**, 151–158 (2002).
37. Veiga, T. *et al.* Metabolic engineering of B-oxidation in *Penicillium chrysogenum* for improved semi-synthetic cephalosporin biosynthesis. *Metab. Eng.* **14**, 437–448 (2012).
38. Yaegashi, J., Oakley, B. R. & Wang, C. C. C. Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in *Aspergillus nidulans*. *J. Ind. Microbiol. Biotechnol.* **41**, 433–442 (2014).
39. Mattern, D. J., Valiante, V., Unkles, S. E. & Brakhage, A. A. Synthetic biology of fungal natural products. *Front. Microbiol.* **6**, 775 (2015).
40. Lubertozzi, D. & Keasling, J. D. Developing *Aspergillus* as a host for heterologous expression. *Biotechnol. Adv.* **27**, 53–75 (2009).
41. Anyaogu, D. C. & Mortensen, U. H. Heterologous production of fungal secondary metabolites in *Aspergilli*. *Front. Microbiol.* **6**, 1–6 (2015).
42. Tsunematsu, Y., Ishiuchi, K., Hotta, K. & Watanabe, K. Yeast-based genome mining, production and mechanistic studies of the biosynthesis of fungal polyketide and peptide natural products. *Nat. Prod. Rep.* **30**, 1139–1149 (2013).
43. Crawford, L. *et al.* Production of cephalosporin intermediates by feeding adipic acid to recombinant *Penicillium chrysogenum* strains expressing ring expansion activity. *Bio-Technology* **13**, 58–62 (1995).
44. Nielsen, M. T. *et al.* Heterologous Reconstitution of the Intact Geodin Gene Cluster in *Aspergillus nidulans* through a Simple and Versatile PCR Based Approach. *PLoS One* **8**, e72871 (2013).
45. Itoh, T. *et al.* Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. *Nat. Chem.* **2**, 858–864 (2010).
46. Sakai, K., Kinoshita, H. & Nihira, T. Heterologous expression system in *Aspergillus oryzae* for fungal biosynthetic gene clusters of secondary metabolites. *Appl. Microbiol. Biotechnol.* **93**, 2011–2022 (2012).
47. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
48. Eldarov, M. A. *et al.* Complete mitochondrial genome of compactin-producing fungus *Penicillium solitum* and comparative analysis of Trichocomaceae mitochondrial genomes. *FEMS Microbiol. Lett.* **329**, 9–17 (2012).
49. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).

50. Zhao, Z., Liu, H., Wang, C. & Xu, J. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**, 1471–2164 (2013).
51. Christensen, T. *et al.* High level expression of recombinant genes in *Aspergillus oryzae*. *Nat. Biotechnol.* **6**, 1419–1422 (1988).
52. Richardson, A. E. & Simpson, R. J. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. *Plant Physiol.* **156**, 989–996 (2011).
53. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, 41 (2013).
54. Numan, M. T. & Bhosle, N. B. Alpha-L-arabinofuranosidases: the potential applications in biotechnology. *J. Ind. Microbiol. Biotechnol.* **33**, 247–260 (2006).
55. Terrasan, C. R. F., Temer, B., Duarte, M. C. T. & Carmona, E. C. Production of xylanolytic enzymes by *Penicillium janczewskii*. *Bioresour. Technol.* **101**, 4139–4143 (2010).
56. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
57. Martinez, D. *et al.* Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* **26**, 553–560 (2008).
58. Regueira, T. B. *et al.* Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. *Appl. Environ. Microbiol.* **77**, 3035–3043 (2011).
59. Horak, R. M., Steyn, P. S., Van Rooyen, P. H., Vleggaar, R. & Rabie, C. J. Structures of the australides A–E, five novel toxic metabolites from *Aspergillus ustus*. *J. Chem. Soc., Chem. Commun.* 1265–1267 (1981).
60. Sobolevskaya, M. P. *et al.* New metabolites from the alga-derived fungi *Penicillium thomii* Maire and *Penicillium lividum* Westling. *Phytochem. Lett.* **15**, 7–12 (2016).
61. Hu, J. *et al.* Characterization of two cytochrome P450 monooxygenase genes of the pyripyropene biosynthetic gene cluster from *Penicillium coprobium*. *J. Antibiot. (Tokyo)*. **64**, 221–227 (2011).
62. Kim, Y. K. *et al.* Pyripyropenes, novel inhibitors of Acyl-CoA - cholesterol acyltransferase produced by *Aspergillus fumigatus*. 2. Structure elucidation of pyripyropene-A, pyripyropene-B, pyripyropene-C and pyripyropene-D. *J. Antibiot. (Tokyo)*. **47**, 154–162 (1994).
63. Wang, H. J., Gloer, J. B., Wicklow, D. T. & Dowd, P. F. Aflavinins and other antiinsectan metabolites from the ascostromata of *Eupenicillium crustaceum* and related species. *Appl. Environ. Microbiol.* **61**, 4429–4435 (1995).
64. Omura, S., Tomoda, H., Kim, Y. K. & Nishida, H. Pyripyropenes, highly potent inhibitors of Acyl-CoA: cholesterol acyltransferase produced by *Aspergillus fumigatus*. *J. Antibiot.* **46**, 1168–1169 (1993).
65. Hayashi, A., Arai, M., Fujita, M. & Kobayashi, M. Pyripyropenes, fungal sesquiterpenes conjugated with alpha-pyrone and pyridine moieties, exhibits anti-angiogenic activity against human umbilical vein endothelial cells. *Biol. Pharm. Bull.* **32**, 1261–1265 (2009).
66. Ariza, M. R., Larsen, T. O., Petersen, B. O., Duus, J. & Barrero, A. F. *Penicillium digitatum* metabolites on synthetic media and citrus fruits. *J. Agric. Food Chem.* **50**, 6361–6365 (2002).
67. Gao, X. *et al.* Fungal indole alkaloid biosynthesis: genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*. *J. Am. Chem. Soc.* **133**, 2729–2741 (2011).
68. Clardy, J., Springer, J. P., Buchi, G., Matsuo, K. & Wightman, R. Tryptoquivaline and Tryptoquivalone, Two Tremorgenic Metabolites of *Aspergillus clavatus*. *J. Am. Chem. Soc.* **97**, 663–665 (1975).
69. Cole, R. J. & Cox, R. H. *Handbook of toxic fungal metabolites*. (Academic Press, New York, 1981).
70. Frisvad, J. C. *et al.* Four psychrotolerant species with high chemical diversity consistently producing cycloaspeptide A, *Penicillium jamesonlandense* sp. nov., *Penicillium ribium* sp. nov., *Penicillium soppii* and *Penicillium lanosum*. *Int. J. Syst. Evol. Microbiol.* **56**, 1427–1437 (2006).
71. Wiemann, P. *et al.* Prototype of an intertwined secondary-metabolite supercluster. *Proc. Natl. Acad. Sci. USA* **110**, 17065–17070 (2013).
72. Eble, T. E. & Hanson, F. R. Fumagillin, an antibiotic from *Aspergillus fumigatus* H-3. *Antibiot. Chemother.* **1**, 54–58 (1951).
73. Fujisawa pharm Co. Ltd. (Fuji-C). Immuno-regulator agents obtained by fermentation of *Penicillium* species. *Japanese patent JP 61257921-A* (1986).
74. Vansteelandt, M. *et al.* Ligerin, an antiproliferative chlorinated sesquiterpenoid from a marine-derived *Penicillium* strain. *J. Nat. Prod.* **76**, 297–301 (2013).
75. Ingber, D. *et al.* Synthetic analogues of fumagillin that inhibit angiogenesis and suppress tumour growth. *Nature* **348**, 555–557 (1990).
76. Bloch, P., Tamm, C., Bollinger, P., Petcher, T. J. & Weber, H. P. Pseurotin, a new metabolite of *Pseudeurotium ovalis* Stolk having an unusual hetero-spirocyclic system. *Helv. Chim. Acta* **59**, 133–137 (1976).
77. Komagata, D., Fujita, S., Yamashita, N., Saito, S. & Morino, T. Novel neuritogenic activities of pseurotin A and penicillic acid. *J. Antibiot. (Tokyo)*. **49**, 958–959 (1996).
78. Ishikawa, M. *et al.* Pseurotin A and its analogues as inhibitors of immunoglobulin E production. *Bioorganic Med. Chem. Lett.* **19**, 1457–1460 (2009).
79. Frandsen, R. J. N. *et al.* The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones. *Mol. Microbiol.* **61**, 1069–1080 (2006).
80. Frandsen, R. J. N. *et al.* Two novel classes of enzymes are required for the biosynthesis of aurofusarin in *Fusarium graminearum*. *J. Biol. Chem.* **286**, 10419–10428 (2011).
81. Machida, M. *et al.* Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**, 1157–1161 (2005).
82. Taniwaki, M. H. *et al.* *Penicillium excelsum* sp. nov. from the Brazil Nut Tree Ecosystem in the Amazon Basin. *PLoS One* **10**, e0143189 (2015).
83. Guo, C.-J. *et al.* Biosynthetic pathway for the epipolythiodioxopiperazine acetylarnotin in *Aspergillus terreus* revealed by genome-based deletion analysis. *J. Am. Chem. Soc.* **135**, 7205–7213 (2013).
84. Simpson, J., Wong, K. & Jackman, S. ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
85. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
86. Vezzi, F., Narzisi, G. & Mishra, B. Reevaluating Assembly Evaluations with Feature Response Curves: GAGE and Assemblathon. *PLoS One* **7**, e52210 (2012).
87. Carson, H. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
88. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
89. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, 445–451 (2012).
90. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
91. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).



92. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
93. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
94. Kildgaard, S. *et al.* Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC-DAD-QTOFMS and a MS/HRMS library. *Mar. Drugs* **12**, 3681–3705 (2014).
95. Lin, H.-C. *et al.* Generation of complexity in fungal terpene biosynthesis: discovery of a multifunctional cytochrome P450 in the fumagillin pathway. *J. Am. Chem. Soc.* **136**, 4426–4436 (2014).

## Acknowledgements

S.G., J.C.N. and M.R. are financially supported by the European Commission Marie Curie Initial Training Network Quantfung (FP7-People-2013-ITN, Grant 607332). The computations were performed on resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC). Sequencing support was provided by the Science for Life Laboratory (SciLifeLab), National Genomics Infrastructure (NGI) and UPPMAX (UPPNEX project ID: b2014081). Support on genome annotation by the National Bioinformatics Infrastructure Sweden (NBIS) is gratefully acknowledged. J.C.F. thanks Agilent for a Agilent Thought Leader Award and a Novo Nordisk Foundation grant (NNF13OC0005201) and the MycoFuelChem project (Grant no. 11-116803). We acknowledge Per Væggemose Nielsen for collection of the soil sample from which the ex type strain of *P. arizonense* was isolated and Martin Kogle for assistance with genomic DNA extraction.

## Author Contributions

S.G. and M.R. performed the cultivations, the chemical extractions and analysis of metabolites and co-wrote the manuscript. J.C.N. performed the genome sequence analysis and co-wrote the manuscript. S.G., K.F.N. and J.C.F. evaluated the data on secondary metabolites. S.G., J.C.N. and J.C.F. described the new species. M.W., J.C.F. and J.N. designed and supervised the work and co-wrote the manuscript.

## Additional Information

**Accession codes:** The *Penicillium arizonense* whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LXJU00000000. The version described in this paper is version LXJU01000000.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Grijseels, S. *et al.* *Penicillium arizonense*, a new, genome sequenced fungal species, reveals a high chemical diversity in secreted metabolites. *Sci. Rep.* **6**, 35112; doi: 10.1038/srep35112 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016