Antonie van Leeuwenhoek (2022) 115:1009–1029
https://doi.org/10.1007/s10482-022-01746-4

ORIGINAL PAPER

Taxonomy, comparative genomics and evolutionary insights of *Penicillium ucsense*: a novel species in series *Oxalica*

Alexandre Rafael Lenz · Eduardo Balbinot · Fernanda Pessi de Abreu · Nikael Souza de Oliveira · Roselei Claudete Fontana · Scheila de Avila e Silva · Myung Soo Park · Young Woon Lim · Jos Houbraken · Marli Camassola · Aldo José Pinheiro Dillon

Received: 2 December 2021 / Accepted: 3 May 2022 / Published online: 9 June 2022

Abstract The genomes of two *Penicillium* strains were sequenced and studied in this study: strain 2HH was isolated from the digestive tract of *Anobium punctatum* beetle larva in 1979 and the cellulase hypersecretory strain S1M29, derived from strain 2HH by a long-term mutagenesis process. With these data, the strains were reclassified and insight is obtained on molecular features related to cellulase hyperproduction and the albino phenotype of the mutant. Both strains were previously identified as *Penicillium echinulatum* and this investigation indicated that these should be reclassified. Phylogenetic and phenotype data showed that these strains represent a new *Penicillium* species in series *Oxalica*, for which the name *Penicillium ucsense* is proposed here. Six additional strains (SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933) collected from the marine environment in the Republic of Korea were also classified as this species, indicating a worldwide distribution of this new taxon. Compared to the closely related strain *Penicillium oxalicum* 114-2, the composition of cell wall-associated proteins of *P. ucsense* 2HH shows five fewer chitinases, considerable differences in the number of proteins related to β-D-glucan metabolism. The genomic comparison of 2HH and S1M29 highlighted single amino-acid substitutions in two major proteins (BGL2 and FlbA) that can be associated with the hyperproduction of cellulases. The study of

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10482-022-01746-4.

A. R. Lenz (✉) · E. Balbinot · F. P. de Abreu · N. S. de Oliveira · S. de Avila e Silva
Bioinformatics and Computational Biology Laboratory, Institute of Biotechnology, University of Caxias Do Sul, Francisco Getúlio Vargas Street 1130, Caxias do Sul, RS 95070-560, Brazil
e-mail: arlenz@ucs.br; alenz@uneb.br

A. R. Lenz
Bahia State University, Silveira Martins Street 2555, Salvador, BA 41150-000, Brazil

R. C. Fontana · M. Camassola · A. J. P. Dillon
Laboratory of Enzymes and Biomass, Institute of Biotechnology, University of Caxias Do Sul, Francisco Getúlio Vargas Street 1130, Caxias do Sul, RS 95070-560, Brazil

M. S. Park · Y. W. Lim
School of Biological Sciences and Institution of Microbiology, Seoul National University, Seoul 08826, South Korea

J. Houbraken
Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

melanin pathways shows that the S1M29 albino phenotype resulted from a single amino-acid substitution in the enzyme ALB1, a precursor of the 1,8-dihydroxynaphthalene (DHN)-melanin biosynthesis. Our study provides important knowledge towards understanding species distribution, molecular mechanisms, melanin production and cell wall biosynthesis of this new *Penicillium* species.

Keywords Albinism · Cellulases · *Penicillium ucsense* · Revised taxonomy · Series *Oxalica* · Filamentous fungi

Introduction

Enzymes are significant for a range of industrial applications, such as paper, food, animal feed, chemical and biofuel production. The degradation of plant polysaccharides by fungal enzymes is widely employed in large-scale industrial processes (Hyde et al. 2019). The screening for enzyme-producing microorganisms, especially cellulose-degrading fungi, has been the subject of research for many decades. In some instances, symbiotic relationships are established between cellulose-degrading microorganisms and organisms that consume plant biomass but lack the machinery to process it (Parkin 1940). One example of this mutualistic exchange are symbiont fungi found in the digestive tract of wood beetle larvae, such as *Anobium punctatum*. Moreover, the breakdown of cellulose in these beetles is achieved through a collective effort between various insect, bacterial and fungal digestive enzymes (Parkin 1940).

In 1979, a *Penicillium* species was isolated from the digestive tract of *A. punctatum* larva, found on a wood wall mural in the Rectory building of the University of Caxias do Sul, Rio Grande do Sul-Brazil. This species was found as a symbiont living in the gut of the beetle larva and the strain was labelled 2HH (Carrau et al. 1981). In the 1990s, this strain was identified using morphology as *Penicillium echinulatum* (unpublished data) and various long-term improvement studies used this name (Camassola et al. 2004; Dillon et al. 2006, 2011; Camassola and Dillon, 2006, 2007, 2009, 2010, 2012; Rubini et al. 2010; Ribeiro et al. 2012; dos Reis et al. 2013; Novello et al. 2014; Schneider et al. 2014, 2016, 2018, 2020; Lenz et al. 2020, 2022). At that time, incorrect strain identification occurred more often, since identification using phenotype is more prone to errors than current molecular techniques. Currently, the taxonomic structure of the genus is well defined and molecular markers facilitate classification and taxonomy (Houbraken et al. 2020).

Long-term improvements with strain 2HH resulted in the mutant S1M29, obtained from the mutant 9A02S1 by employing hydrogen peroxide mutagenesis and a selection of mutants in a medium supplemented with 2-deoxyglucose (Dillon et al. 2011). The S1M29 mutant provides a better biomass hydrolysis and is the best mutant to date, due to a significant increase in enzyme titers as a result of several accumulated mutations (Schneider et al. 2016, 2018). Moreover, the production of white-coloured conidia by S1M29 is another phenotypic difference between the wild-type and the mutant.

On the whole, the cell wall is potentially the part of the cell that exhibits the most phenotypic diversity and plasticity, comprising an attractive set of elements to explore genome evolution. Fungal cell walls are dynamic structures that play a critical role in fungal survival, growth, and morphology. Its major constituents are chitin, chitosan, β-1,3-glucan, β-1,6-glucan, mixed β-1,3-/β-1,4-glucan, α-1,3-glucan, glycoproteins and melanin. In general, the fungal cell wall is generated by the cross-linking of glucans, chitin, melanin and other cell wall proteins to create a three-dimensional matrix (Free 2013).

Melanin pigments are formed by oxidative polymerization of phenolic compounds. Many fungi synthesise melanins, and several types of melanin are known to exist in the fungal kingdom (Eisenman and Casadevall 2012). Melanin provides defence against environmental stresses such as ultraviolet light, oxidising agents and ionising radiation. In addition, it contributes to the ability of fungi to survive in harsh environments. Also, fungal melanin contributes to virulence in an array of human pathogen fungi, including *Cryptococcus neoformans*, *Paracoccidioides brasiliensis*, *Aspergillus fumigatus* and *Talaromyces marneffei* (Langfelder et al. 2003; Eisenman and Casadevall 2012).

Fungi may produce melanin via distinct pathways: the eumelanin via the 1,8-dihydroxynaphthalene (DHN) and L-3,4-dihydroxyphenylalanine (DOPA) pathways, and the pyomelanin via L-tyrosine degradation pathway. DHN-melanin, responsible for dark

polymer production, is probably the best characterised fungal melanin biosynthetic pathway. Melanin biosynthesis homologues of these three pathways have been characterised in several filamentous fungi. However, chemical characterization of melanin can be a challenging task as the pigment is highly heterogeneous, insoluble in organic solvents, hydrophobic, and resistant to chemical degradation (Pralea et al. 2019).

The main goals in this study are (1) to reclassify and describe *Penicillium* strain 2HH using morphology and multigene phylogeny, (2) to publish the draft genome of this new *Penicillium* species, (3) to elucidate molecular features related to cellulase hyperproduction and albinism, and (4) to get insight in the genome evolution of this species, considering that genome information from non-model organisms is highly important, as they represent specific phenotypes that aid in disengaging the common parts of gene sets from those that have evolved as adaptations to specific ecosystems.

Materials and methods

Strains

The studied wild-type strain was isolated in 1979 from the digestive tract of *Coleoptera* larva in Caxias do Sul, Brazil and deposited in the culture collection of the Laboratory of Enzymes and Biomass (University of Caxias do Sul; Caxias do Sul, Brazil) under accession number 2HH. This wild-type strain was also deposited in the CBS culture collection housed in the Westerdijk Fungal Biodiversity Institute (Utrecht, the Netherlands) under accession number CBS 146492 in 2020 and the Culture Collection of Filamentous Fungi (CCFF) housed in the Oswaldo Cruz Foundation (Rio de Janeiro, Brazil) under accession number IOC 4717 in 2022. The studied mutant strain is also deposited in the culture collection of the Laboratory of Enzymes and Biomass (University of Caxias do Sul; Caxias do Sul, Brazil) under accession number S1M29. Strains SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933 are deposited at the Seoul National University Fungus Collection (SFC), Republic of Korea.

Morphology

Morphological examination of the strain was performed using the procedures published in Visagie et al. (2014). In short, macroscopic characters were recorded from colonies grown for 7 days at 25 °C on Czapek yeast extract agar (CYA), dichloran 18% glycerol agar (DG18), malt extract agar (MEA, Oxoid®), yeast extract sucrose agar (YES) and creatine agar (CREA). Additional CYA plates were incubated at 15, 30 and 37 °C. Microscopic observations were made from colonies grown on MEA and slides were made in lactic acid (60%). A Nikon SMZ25 dissecting microscope and a Zeiss Axio Imager.A2 with Differential Interference Contrast (DIC) microscope, both equipped with Nikon DS-Ri2 cameras, were used to take photos of the studied strain. The potential formation of a sexual state was studied by growing the strain for 40 days at 25 °C on MEA, CYA and oatmeal agar (OA, Traditional Rolled Oats; Quaker Oats).

Phylogenetic analysis

The phylogenetic relationships were studied using four single locus sequence datasets and a combined dataset of all four loci. The single locus datasets contained sequence data of the following molecular markers: the 5.8S rDNA subunit including the internal transcribed spacer regions (ITS), β-tubulin (*BenA*), calmodulin (*CaM*) and RNA polymerase II second largest subunit (*RPB2*). The complete sequence of each marker was identified and annotated in the whole genome sequence of strain 2HH. These annotated sequences were included in the whole genome sequence deposited at GenBank.

Each single locus sequence dataset was compiled by combining the newly obtained sequences with publicly available (reference) sequences of phylogenetically related *Penicillium* species (Kubátová et al. 2019; Houbraken et al. 2020). Strains and respective sequence accession numbers are listed in Suppl. Table S01. Alignment of each sequence dataset was performed using MAFFT (v. 7.480) (Katoh et al. 2018) and manually optimised using AliView (v. 1.26) (Larsson 2014). Subsequently, JModelTest2 (v. 2.1.10) (Darriba et al. 2012) was used to find the best substitution model, which was selected by the corrected Akaike Information Criterion (AICc) for

each dataset. The resulting alignments and models (Table 1) were used for Maximum Likelihood (ML) and Bayesian analyses.

The concatenated dataset was obtained by FaBox (v. 1.5) (Villesen 2007) and missing sequences in the *CaM* and *RPB2* datasets were replaced by ?'s to indicate missing data. MrBayes (v. 3.2.7a) (Ronquist et al. 2012) was used for the Bayesian analysis using the following settings: GTR (nst = 6) + GAMMA, 10^7 generations, sampling every 1000 generations with a burnin fraction of 0.25. The maximum likelihood analysis was performed using the RAxML-HPC2 (v. 8.2.12) (Stamatakis 2014) using the settings GTR + GAMMA, executing ML search and thereafter a thorough bootstrap with 1,000 iterations. The four single loci were defined for the concatenated dataset. The CIPRES Science Gateway (v. 3.3) (Miller et al. 2010) was used to perform both analyses.

Trees were visualised in FigTree (v. 1.4.4) (Rambaut 2009) and edited in Inkscape (v. 0.92.2) (Inkscape 2019). Posterior probabilities (pp) values and bootstrap support (bs) values are labelled at the top of a branch. Values less than 0.95 pp and 80% bootstrap support are not shown and indicated with a hyphen, and branches with full support are thickened. *Penicillium echinulatum* CBS 317.48 (sect. *Fasciculata*) was chosen as an outgroup.

DNA and RNA sequencing

Strains

Both strains (wild-type 2HH and mutant S1M29) were used for DNA sequencing, while only the mutant S1M29 was used for RNA sequencing.

Table 1 Sequence data sets and models used in the phylogenetic analyses

Dataset	bp	Taxa	Substitution model
ITS	584	132	K80 + I + G
BenA	512	52	TIM2ef + G
CaM	577	37	TrNef + I + G
RPB2	1050	34	TrN + G
Combined dataset	2719	52	TIM2 + I + G (ITS) TPM3uf + I + G (*BenA*) TrNef + I + G (*CaM*) TrN + G (*RPB2*)

Growth conditions

Both strains were grown for DNA isolation in 100 mL of cellulose agar (agar-C) consisting of 40 mL of swollen cellulose, 10 mL of mineral solution (KH_2PO_4, 10 g; $MgSO_4.7H_2O$, 6 g; $CO(NH_2)_2$, 6 g; $CaCl_2$, 6 g; $FeSO_4.7H_2O$, 0.1 g; $MnSO_4.H_2O$, 0.0312 g); $ZnSO_4.7H_2O$, 0.028 g; $CoCl_2.6H_2O$, 0.04 g; 1 L distilled water), 0.1 g of proteose peptone (Oxoid L85®), 2 g of agar, and 50 mL of distilled water. The strains were grown in inclined tubes on C-agar for 7 days at 28 °C until the formation of conidia.

Mutant strain S1M29 was grown for RNA isolation in a medium containing 0.2% (w/v) peptone; 0.05% (w/v) Prodex®; 1% (w/v) sugar cane bagasse pretreated by steam explosion (Usina Vale do Rosário, Morro Agudo, SP, Brazil) as carbon source; 0.1% (v/v) Tween 80®; 0.002% (v/v) antibiotic ciprofloxacin (Proflox®, EMS S/A); 5% (v/v) mineral solution (described above); and distilled water to complete a final volume of 100 mL. Erlenmeyer flasks (500 mL), containing 100 mL of the production medium, were inoculated with a suspension of 1×10^5 conidia per mL and kept at 28 °C, in a reciprocal agitation of 180 rpm for 120 h.

DNA and RNA isolation

Isolation of high-molecular-weight genomic DNA from the 2HH wild-type and mutant S1M29 strains was performed using the protocol outlined in Green and Sambrook (2012). Isolation of RNA from the mutant S1M29 was performed using TRIzol® Reagent, following the manufacturer's protocols (Chomczynski 1993). Then, RNA purification was performed using TURBO™ DNase.

Sequencing

DNA samples were prepared using the Illumina protocol outlined in "TruSeq DNA Sample Preparation Guide" (Part# 15,005,180 Rev. A November 2010), generating libraries for Illumina Sequence by Synthesis. RNA sample was prepared using the Illumina protocol outlined in "TruSeq RNA Sample Preparation Guide" (Part# 15,008,136 Rev. A November 2010), generating non-stranded RNA libraries with poly-A selection.

The sequencing was conducted by the commercial provider Ambry Genetics (Aliso Viejo, CA, USA) in 2013, performing DNA sequencing of two samples (2HH and S1M29) and RNA sequencing of one sample (S1M29). Ambry Genetics performed Quality Control, prepared mRNA and DNA libraries, and sequenced on three lanes of the Illumina HiSeq 2000, using 100 bp paired-end processing.

Genome assemblies and quality assessment

Raw data quality control (QC) assessment using FastQC (v0.11.5) (Andrews et al. 2015) was performed to check the overall sequence quality, the GC percentage distribution and the presence/absence of overrepresented sequences. Using Trim Galore (v0.4.4) (Krueger 2015), small fragments (length, 30 bp) were excluded, following adapter clipping and quality trimming. Low quality bases from the 3' and 5' ends were removed before being cut and quality trimmed (sequencing quality values, Q28).

The high-quality reads were assessed using KmerGenie (v1.7044) (Chikhi and Medvedev 2014), to predict the best k-mer and genome size. The high-quality reads were assembled with SPAdes (v3.11.0) (Bankevich et al. 2012). SPAdes is a multi-k-mer assembler, we used odd values ($21 \leq k \leq 83$). Further, scaffolds shorter than 500 bp were removed. Quast (v5.0.2) (Gurevich et al. 2013) with default parameters was used to statistically evaluate the assemblies. BUSCO (v3.0.2) (Waterhouse et al. 2018) assembly mode was used to assess the assemblies, providing quantitative measures based on evolutionarily expectations of gene content from near-universal single-copy orthologous selected from OrthoDB (v9) (Kriventseva et al. 2015). Five conserved orthologous datasets were used for evaluation: *Eukaryota* (303 genes), *Fungi* (290 genes), *Dikarya* (1312 genes), *Ascomycota* (1315 genes) and *Eurotiomycetes* (4046 genes).

Transcriptome assembly for gene calling

Reads assessment and quality trimming were performed in the same way as genome sequences. Paired-end reads were aligned over both genomes using hisat2 (v2.0.5) (Kim et al. 2015), and assembled into transcripts using Trinity (v2.5.1) (Grabherr et al. 2011). In addition, Trinity de novo transcriptome assembly was performed. All transcripts were used to train gene finders for gene model prediction.

Gene calling

Repeat sequences of both assemblies were masked using RepeatMasker (v4.0.7_2) (Smit et al. 2013), RepeatModeler (v1.0.8_1) (Smit et al. 2013) and RepBase library (2017-01-27) (Bao et al. 2015). Repeat masked assemblies were used for coding gene prediction, performed independently with a set of gene finders. The first group uses an *ab-initio* approach to predict genes directly from nucleotide sequences, including Augustus (v3.2.2) (Stanke et al. 2006) and GeneMark-ES (v4.33) (Ter-Hovhannisyan et al. 2008). Augustus parameters were trained on gene models of the S1M29 mutant with the transcriptome data as hints. Pre-trained parameters from *Aspergillus* (*A. fumigatus*, *A. nidulans*, *A. oryzae* and *A. terreus*) were also used in Augustus predictions.

The second group uses a similarity-based approach to identify gene structure using a sequence alignment between genomic sequence and transcript or protein databases. BLAT (v36) (Kent 2002) and GMAP-GSNAP (v2017-06-20) (Wu et al. 2016) were used to align the RNA-Seq transcripts of the S1M29 mutant, while Exonerate (v2.2.0) (Slater and Birney 2005) was used to align UniProtKB/Swiss-Prot (The UniProt Consortium 2019) and the proteome of *P. oxalicum* 114-2 (Liu et al. 2013). Additionally, tRNAscan-SE (v2.0.3) (Lowe and Chan 2016) was used to predict transfer RNAs (tRNAs). EVidence Modeler (v0.1.3) (Haas et al. 2008) was used to take all gene prediction inputs, outputting consensus gene models and generating the predicted gene set for each strain. *Penicillium* spp. was used to revise and complement the predicted genes by homology searches using Exonerate (Slater and Birney 2005) and SoftBerry Fgenesh + (Solovyev 2008). The annotation completeness was assessed by running BUSCO (v3.0.2) (Waterhouse et al. 2018) in protein mode, using the same datasets as the assembly assessment.

Functional annotation

Predicted proteins were functionally annotated using the standard protocol (McDonnell et al. 2018), implemented by the tool Seq2Annot. All predicted gene models were functionally annotated using

Seq2Annot workflow: SignalP Server (v5.0) (Almagro Armenteros et al. 2019) to predict the presence and location of signal peptide cleavage sites; TMHMM Server (v2.0) (Moller et al. 2002) was used to identify transmembrane helices and membrane-bounds; InterProScan (v5.25–64.0) (Jones et al. 2014) was used to map Interpro families, domains and gene ontology (GO) terms; hmmscan (v3.1b2) (Johnson et al. 2010) was used to identify PFAM domains over PFAM database (v31.0-2017-02) (El-Gebali et al. 2019) using gathering cutoffs; KO (KEGG Orthology (Kanehisa et al. 2016)) assignments and automatically generated KEGG pathways were assigned to predicted proteins using KAAS (Moriya et al. 2007); KEGG hints were also used to assign Enzyme Comission (EC) numbers; EggNOG-Mapper (v2) (Huerta-Cepas et al. 2017) was used to map general functional categories from Clusters of Orthologous Groups (COGs) and orthologous precomputed eggNOG clusters from the eggNOG (v5.0) (Huerta-Cepas et al. 2019) (ascNOG) database; secondary metabolite biosynthetic gene clusters were predicted by antiSMASH fungal (v3.0) (Blin et al. 2017), and mibig database (v1.3) (Epstein et al. 2018).

Moreover, product assignment of Carbohydrate-Active Enzymes (CAZymes), sugar transporters and transcription factors were analysed manually, using orthologous from related fungi, BLASTp (v2.7.1 +) (Camacho et al. 2009) searches over UniProtKB/Swiss-Prot/TrEMBL (The UniProt Consortium 2019), and specific databases, as described below. Two approaches were combined in order to improve CAZyome annotation accuracy: (i) running dbCAN2 (Zhang et al. 2018) server upon HMMdb release 8.0 and (ii) running BLASTp (v2.7.1 +) (Camacho et al. 2009) upon CAZyDB (07/31/2019).

To assess the diversity of sugar transporters, the domain PF00083 (08/05/2018) profile extracted from the PFAM database (El-Gebali et al. 2019) was used to search over *P. ucsense* proteomes with hmmsearch (v3.1b2) (Johnson et al. 2010), choosing hmmsearch score $\geq$ 238 as cutoff (Peng et al. 2018). This class of sugar transporters was chosen for its biotechnological relevance, related to the production of cellulolytic enzymes. To assess the diversity of transcription factors, Interpro and PFAM domains of whole proteomes were used to search TF domains using those protein domains described in the library of TFs used by CIS-BP Database (Weirauch et al. 2014).

Investigation of cell wall-associated proteins

Previously released cell wall-associated proteins from *A. fumigatus* Af293 (Bernard and Latgé 2001; Free 2013; Teixeira et al. 2017) and *N. crassa* OR74A (Patel and Free 2019) were used to find orthologous cell wall-related proteins in *P. ucsense* 2HH and *P. oxalicum* 114-2. Proteomes obtained from UniProtKB (The UniProt Consortium 2019) were used to find orthologous groups in whole genome-wide searches using ProteinOrtho (V5.16b) (Lechner et al. 2011).

Origin tracking of cellulase hyperproduction

PogressiveMauve (v2015-02-13) build 0 (Darling et al. 2010) was used to align both *P. ucsense* draft genomes. Protein sequences of both proteomes (2HH and S1M29) were aligned using MAFFT online service (v7.452) (Katoh et al. 2018), in order to find amino-acid substitutions generated by the mutations. Proteins that contained amino-acid substitutions in the S1M29 mutant were classified in four levels of probable impact on cellulase hyperproduction; whose classification was previously supported by experts in fungal metabolism. The main proteins affected by the SNPs, classified at levels 3 and 2, exhibit potential involvement in the expression of cellulase coding genes. These levels include transporter proteins, transcription factors and other proteins strongly related to gene regulation. In contrast, levels 1 and 0 comprise a diverse series of proteins with low probability of impact in the expression of cellulolytic enzymes.

Origin tracking of albinism

Previously released melanin-associated proteins from *A. fumigatus* Af293, *Aspergillus niger* CBS 513.88, *Penicillium chrysogenum* P2niaD18 (Teixeira et al. 2017) and *T. marneffei* ATCC 18224 (Woo et al. 2010) were used to identify orthologous melanin-related proteins in *P. ucsense* 2HH and *P. oxalicum* 114-2. Proteomes obtained from UniProtKB (The UniProt Consortium 2019) were used to find orthologous groups in whole genome-wide searches using ProteinOrtho (V5.16b) (Lechner et al. 2011).

Macromorphology and other phenotypic characters were observed in four culture media: (i) MEA; (ii) Blakeslee's Malt extract agar (MEAbl); (iii) CYA; and (iv) In-house Maintenance Medium (IHMM).

The first three culture media are described in Visagie et al. (2014), and the last culture medium is described in Dillon et al. (2006). Cultures were inoculated using a concentrated conidium suspension of both strains 2HH and S1M29 on the four media. After 24 h, the inocula was transferred using a three-point method to the same medium in 9 cm glass Petri dishes and incubated in the dark at 28 °C. The cultures were examined after 7 and 14 days of growth and photos of the dishes were taken and edited in GIMP (v2.10.14) (The GIMP 2020).

Results and discussion

Morphology and phylogenetic analysis

Based on the morphological characterization carried out in the 1990's (unpublished data), strain 2HH was originally identified as *P. echinulatum*. This incorrect identification was likely based on the production of distinctly roughened conidia. After preliminary analysis of the genome sequence, this identification was questionable and therefore the strain was re-identified using the current classification schemes.

To determine the phylogenetic position of strain 2HH, datasets containing *BenA*, *CaM*, ITS and *RPB2* reference sequences from GenBank were compiled, aligned and analysed (Suppl. Table S01). Phylogenetic analyses were performed including individual and concatenated datasets. Table 1 provides an overview of each dataset, including its length and its best substitution model. The multigene phylogram is shown in Fig. 1 and the single gene trees in supplementary data (Suppl. Figs. 1, 2, 3, 4).

Phylogenetic analysis using ITS sequences (Suppl. Figure 1) showed, with high statistical support that strain 2HH belongs to section *Lanata-Divaricata* and series *Oxalica* (Houbraken et al. 2020). This strain resides in well-supported clade (1.00 pp, 86% bs) with strains of the recently described species *Penicillium diatomitis* (CBS 140107 ^{T}, CCF 4379, CCF 3779, MH285, MH272) (Kubátová et al. 2019). Besides these strains, also seven other strains (SFC101850 (Park et al. 2018), SFCP10873, SFCP10886, SFCP10931, SFCP10932, SFCP10933 and 3.16039) cluster in this clade.

Phylogenetic analysis using *BenA* sequences (Suppl. Fig. 2) showed that strains 2HH, SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932, SFCP10933 form a well-supported clade, sister to strains identified as *P. diatomitis* and *P. soosanum*. Three well-supported lineages are present within this clade, one containing the type of *P. diatomitis* (CBS 140107) (1.00 pp, 94% bs), one the type of *P. soosanum* (CBS 140106) (0.96 pp, 98% bs) and one containing strains 2HH, SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933 (fully supported).

Compared to the *BenA* dataset, less *CaM* sequences are available of strains belonging to the *P. diatomitis*/*P. soosanum* clades. Phylogenetic analysis of the *CaM* dataset (Suppl. Fig. 3) shows a similar topology as present in the *BenA* phylogram. Strains 2HH, SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933 form a distinct and well-supported clade (0.99 pp, 98% bs). A similar result is obtained in the phylogenetic analysis of the *RPB2* dataset (Suppl. Fig. 4). The topology of this phylogram is the same as in the *BenA* phylogram. The main difference is the lack of *RPB2* sequences of strains SFCP10873, SFCP10931 and SFCP10933. Even so, the clade containing strains 2HH, SFC101850, SFCP10886 and SFCP10932 is well-supported (0.99 pp, 99% bs). The lineages containing the types of *P. diatomitis* and *P. soosanum* are also well-supported.

The phylogenetic analysis of the combined dataset (Fig. 1) revealed the presence of two fully supported clades in *Penicillium* section *Lanata-Divaricata* series *Oxalica*. One clade comprises strains designated as *P. oxalicum*, including the ex-type strain of *P. oxalicum* CBS 219.30. This clade also includes CBS 173.81, the type of *P. asturianum*, which is considered a synonym of *P. oxalicum*. The other clade contains three lineages and these lineages are also present in the *BenA*, *CaM* and *RPB2* phylogenies. The *P. diatomitis* lineage is supported with 1.00 pp and 99% bs and the *P. soosanum* lineage is supported with 1.00 pp and 95% bs. Lastly, the lineage containing strains 2HH, SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933 is fully supported.

In recent years, a new measure is being adopted by fungal taxonomists: whole-genome Average Nucleotide Identity (ANI) (Gostinčar 2020). We analysed both the whole genome sequences of strain 2HH (GCA_014839855.1) and *P. oxalicum* 114-2 (GCA_000346795.1) obtained from GenBank,

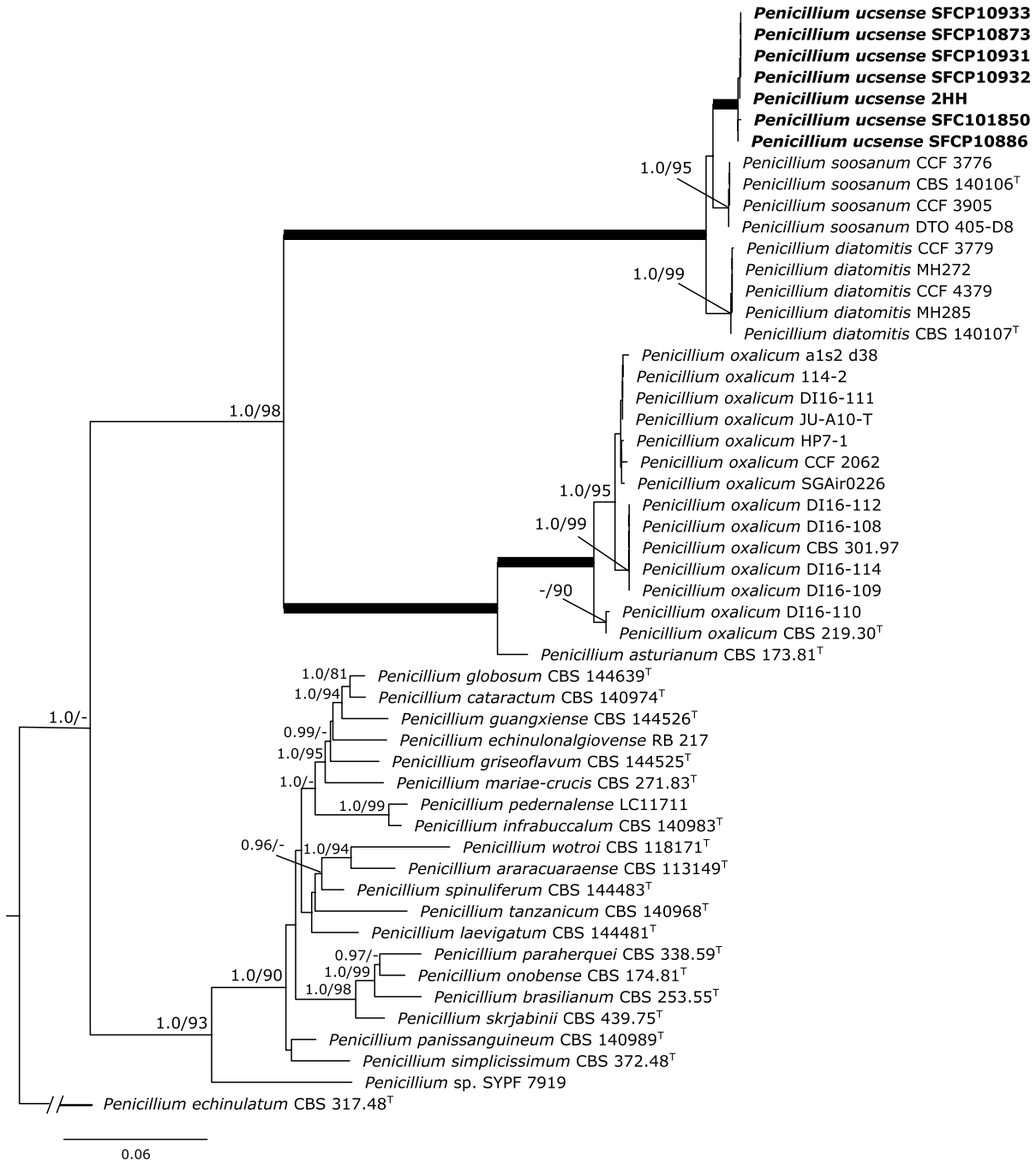


Fig. 1 Phylogenetic tree based on a combined dataset containing ITS, *BenA*, *CaM* and *RPB2* sequences. The phylogram is rooted with *Penicillium echinulatum* CBS 317.48

taking into account that *P. oxalicum* is the closest species that has a whole genome sequenced. We used FastANI (v.0.1.3) to analyse the whole genome sequences of both fungi and the ANI generated by the tool was 79.3%. This result is in agreement with the literature that suggests ANI values below 95% to delineate a new species (Gostinčar 2020).

Unfortunately, there are no whole sequenced genomes of the closest species *P. diatomitis* and *P. soosanum*.

Lastly, strain 2HH was originally identified using morphological features as *P. echinulatum*. However, *P. echinulatum* is classified in the section *Fasciculata* (subg. *Penicillium*) and our phylogenetic analysis evidence that *P. echinulatum* is phylogenetically unrelated to strain 2HH (subg. *Aspergilloides*, sect. *Lanata-Divaricata*).

Species description

Phylogenetic analysis shows that strains 2HH, SFC101850, SFCP10873, SFCP10886, SFCP10931, SFCP10932 and SFCP10933 forms a distinct sister lineage to *P. diatomitis* and *P. soosanum.* The phylogenetic position of this lineage is concordant between the *BenA*, *CaM* and *RPB2* datasets. The strain 2HH is also phenotypically distinct from these species (see *Notes* below) and we therefore describe this strain as a new species named *Penicillium ucsense*.

Penicillium ucsense A. Lenz & Houbraken, *sp. nov.* MycoBank MB 839721. Figure 2.

Etymology: Named after UCS, the abbreviation of University of Caxias do Sul, considering that the fungus was isolated from the digestive tract of *Anobium punctatum* larva that was collected in the Rectory building of this University in 1979 and where it has been studied over the last 40 years.

Type: Brazil, intestinal tract of *Anobium punctatum* larva, 1979, collected and isolated by J.L. Carrau & R.T. Ribeiro. (holotype CBS H-24331, culture ex-type: CBS 146492 = DTO 426-B1 = IOC 4717 = 2HH).

In: *Penicillium* subgenus *Aspergilloides*, section *Lanata-Divaricata*, series *Oxalica*.

ITS barcode: OM914583 (ITS1, 5.8S rDNA & ITS2). Alternative identification markers: *BenA* = ON024157, *CaM* = ON024158, *RPB2* = ON024159. Partial sequences were deposited at GenBank according to the referenced accession numbers. Complete sequences are also included in the WGS deposited at GenBank under accession WIWU00000000.

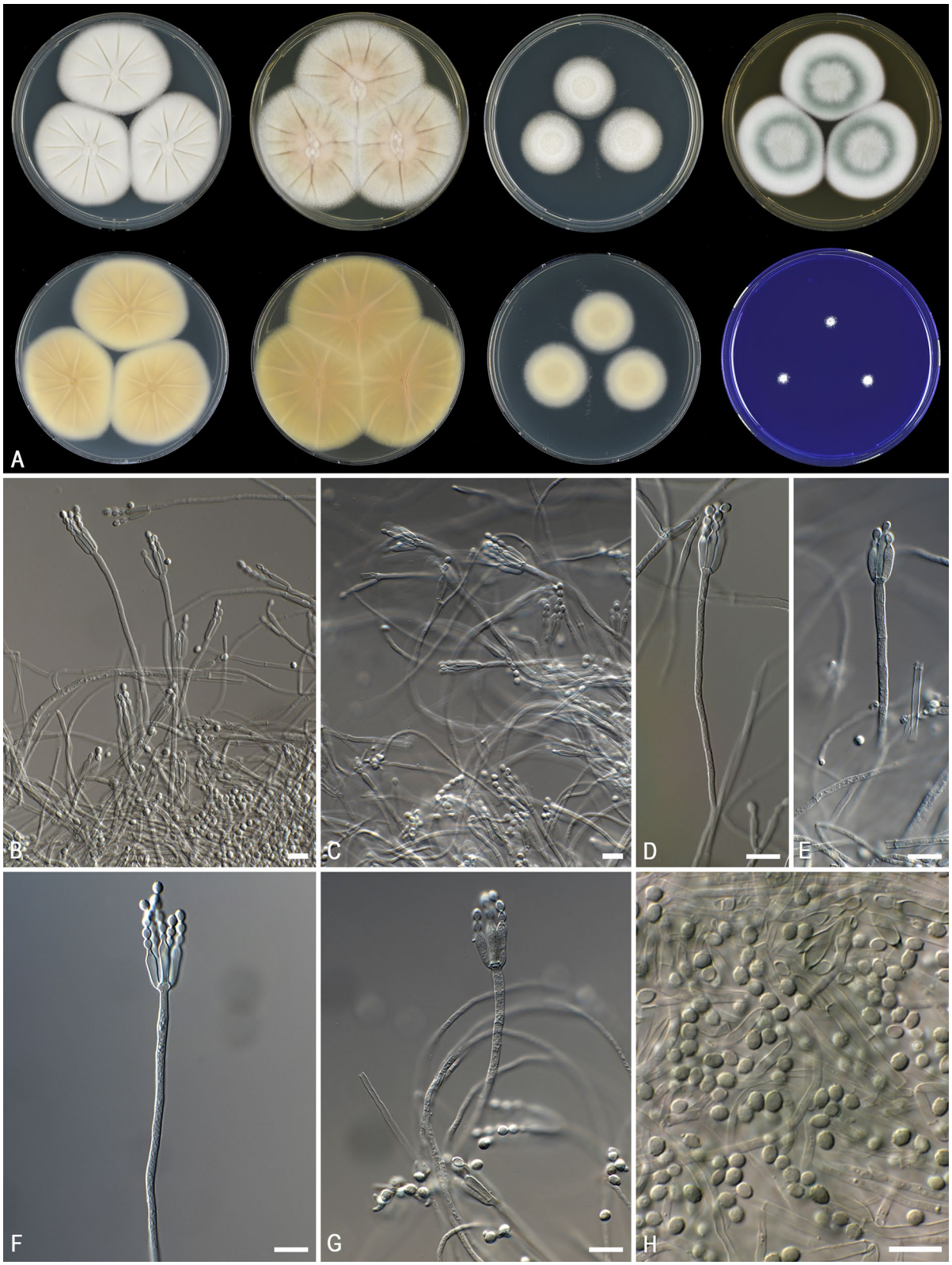
Colony diam (7 d, in mm): CYA 57–60; CYA 15 °C 19–20; CYA 30 °C 58–61; CYA 37 °C no growth; DG18 26–30; MEA 45–50; YES 55–59; CREA 18–22.

Colony characters (7 d, 25 °C): CYA colonies radially sulcate, slightly raised in the centre; margin entire; mycelium white; sporulation absent; soluble pigments absent; exudates absent; conidial colour not determined; reverse pale yellow–brown. *YES*, colonies radially sulcate, slightly elevated; margin entire, slightly feathery; mycelium pale brown; texture velvety; sporulation absent; soluble pigments absent; exudates absent; reverse pale brown. *MEA*, colonies slightly radially sulcate in centre, moderately high; margin entire; mycelium white; colony texture floccose; sporulation absent at the edge, poor in centre and good in a ring between centre and edge; soluble pigments absent; exudates absent; conidial colour *en masse* dull green; reverse unchanged in centre, light brown at edge. *DG18*, colonies plane, raised at centre; margin entire; mycelium white; texture velvety; sporulation absent; soluble pigments absent; exudates absent; reverse brownish white in centre, uncoloured at edge. *CREA*, poor growth, acid and base production absent.

Micromorphology: Ascomata and sclerotia not observed. Conidiophores (60–)75–150 µm long, 2–3 µm in width, smooth, thin, predominantly monoverticillate, rarely biverticillate, non-vesiculate; metulae appressed, in terminal whorl of 1–3, equal in length, (10–)13–16 µm; phialides cylindrical, 2–5 per stipe or metula, 11–14 × 2.5–3 µm; conidia broadly ellipsoidal, rough walled, 3.5–4 × 2.5–3 µm.

Distribution: To date, seven collections are reported of this species. Strain 2HH was isolated in 1979 from the digestive tract of *Anobium punctatum* larva in Caxias do Sul, Brazil; strain SFC101850 (Park et al. 2018) was isolated in 2016 from *Arctoscopus japonicus* (sailfin sandfish) egg masses in Republic of Korea; strains SFCP10873, SFCP10932 and SFCP10933 were isolated in 2017 from sea sand in Republic of Korea; and strains SFCP10886 and SFCP10931 were isolated in 2017 from seaweed in Republic of Korea.

Notes: *Penicillium ucsense* is phylogenetically related to *P. diatomitis*, *P. oxalicum* and *P. soosanum*. These species share a fast growth rate on CYA, YES and MEA and the production of cylindrical phialides and (finely) roughened conidia. However, there are several characters that distinguish *P. ucsense* from these species. The conidiophores of *P. ucsense* are predominantly monoverticillate and sporulation is absent or poor on CYA, DG18, MEA and YES. In


A
B
C
D
E
F
G
H

◂**Fig. 2** Macro- and micromorphological observations of *Penicillium ucsense* 2HH. **a** Colonies: top row left to right, obverse CYA, YES, DG18 and MEA; bottom row left to right, reverse CYA, reverse YES, reverse DG18 and CREA. **b–g** Conidiophores. **h** Conidia. Scale bars: **b–h** = 10 µ

contrast, the other species of this series are predominantly biverticillate branched and sporulation is profuse. Furthermore, *P. oxalicum* (10–36 mm), *P. diatomitis* (7–14 mm) and *P. soosanum* (5–20 mm) are able to grow on CYA incubated at 37 °C, while no growth was observed in *P. ucsense* after 7 days. Furthermore, *P. diatomitis* and *P. oxalicum* produce finely roughened, ellipsoidal conidia, while *P. ucsense* and *P. soosanum* produce broadly ellipsoidal to subglobose, roughened conidia. *Penicillium ucsense* shares the inability to produce acid compounds on CREA with *P. diatomitis*, which sets it apart from the acid producing species *P. oxalicum* and *P. soosanum*.

Sporulation is absent on most media and only observed on MEA. *Penicillium* strains can lose typical characteristics when maintained in culture collections over longer periods. Probably the lack of sporulation and the reduced conidiophore branching in strain 2HH is caused by the long preservation (> 40 years), considering that sporulation was observed in the strains collected from the marine environment in the Republic of Korea. We suggest a new study to morphologically characterise one of the strains collected in the Republic of Korea, in order to complement the knowledge about this novel species.

General features of genomes

The genomes of *P. ucsense* 2HH and S1M29 were sequenced using the Illumina HiSeq 2000 platform and 29,316,764,800 bp (raw coverage of 961×) for the S1M29 mutant and 25,514,587,400 bp (raw coverage of 836×) for the 2HH wild-type were obtained. The high coverage was required to precisely identify the single-nucleotide polymorphisms (SNPs) that occurred during the S1M29 mutagenesis process. These sequencing data were used in previous studies (Schneider et al. 2016; Lenz et al. 2020) and here we publish the Sequence Read Archives (SRA) and WGS of both strains to the scientific community.

The genome assembly of the 2HH wild-type comprises 697 scaffolds, totalizing 30.43 Mb and scaffold N50 equals 151.4 kb. While the genome assembly of the S1M29 mutant comprises 673 scaffolds, totalizing 30.41 Mb, and scaffold N50 is 185.175 kb. Furthermore, the GC contents of both assemblies are 50.3% and unclosed gap regions represent less than 2% for each strain assembly as assessed by Quast. The summary results are presented in Table 2. WGS project of strain 2HH was deposited at DDBJ/ENA/GenBank under the accession WIWU00000000 and the version described in this paper corresponds to WIWU01000000. The WGS project of strain S1M29 is deposited at DDBJ/ENA/GenBank under the accession WIWV00000000 and the version described in this paper corresponds to WIWV01000000.

We evaluate both assemblies using five BUSCO conserved orthologous datasets. The completeness average of the assemblies is higher than 98%, where 1.2% of the core orthologous are missing. The proportion of fragmented orthologous comprises less than 0.6%. Next, we performed gene calling independently with a set of gene finders, resulting in 8366 gene models predicted for 2HH, while 8375 gene models were predicted for S1M29 (Table 2). The annotation completeness was assessed by running BUSCO in protein mode, using the same datasets of the assembly assessment. The average of complete proteins is 99.4%, where just 36 of 7266 orthologous are missing in each strain. The proportion of fragmented proteins comprises less than 0.4%.

Previous proteome completeness of 28 fungal species showed that the average of complete proteins is higher than 95% in most species. High numbers of fragmented proteins (more than 150 in some species) were observed regardless of the genus, technology, assembler or automatic gene calling methodology (Aguilar-Pontes et al. 2018). Our draft assembly evaluation showed results above this average, allowing a plentiful gene calling and comparison with other fungi.

Investigation of cell wall-associated proteins

The overwhelming diversity in cell wall structure combined with variant and recurrent features comprise an attractive set of elements to explore genome evolution. Comparative analyses between previously released cell wall-associated proteins from the well-characterised fungi *A. fumigatus* Af293 (Bernard and Latgé 2001; Free 2013; Teixeira et al. 2017) and *N. crassa* OR74A (Patel and Free 2019) and our dataset

Table 2 Genome features of *Penicillium ucsense* genomes

Features (# means the number)	2HH	S1M29
Raw reads (bp)	255,145,874	293,167,648
Paired-end read length (bp)	100	100
Filtered reads (bp)	239,270,048	273,856,380
Genome assembly size (Mb)	30.43	30.41
# of scaffolds	697	673
GC content (%)	50.3	50.3
# of N's	533	621
Scaffold N50	151,400	185,175
Scaffold N75	86,997	98,992
Scaffold L50	62	50
Scaffold L75	126	109
Average Scaffold (bp)	43,655	45,180
Largest Scaffold (bp)	511,359	802,811
# of gene models	8366	8375
# of tRNAs	188	197
# of putative proteins	8173	8173
Average of exons per gene	3.2	3.2
Average of introns per gene	2.7	2.7
Smallest protein length (aa)	51	51
Average protein length (aa)	519.2	519.2
Largest protein length (aa)	7243	7243
# of Complete and single-copy BUSCOs of *Eurotiomycetes* class	3962	3963
# of Complete and duplicated BUSCOs of *Eurotiomycetes* class	10	10
# of Fragmented BUSCOs of *Eurotiomycetes* class	40	39
# of Missing BUSCOs of *Eurotiomycetes* class	34	34
BUSCO Completeness (%) of *Eurotiomycetes* class	98.1	98.1
# of PFAM annotations	9505	9503
# of InterPro annotations	22,602	22,591
# of GO terms annotations	15,103	15,099
# of Signal peptide annotations	652	651
# of Transmembrane annotations	1666	1668
# of BUSCO annotated (All datasets)	7241	7245
# of EGGNOG annotated	7982	7982
# of COG annotated	7528	7527
# of KEGG annotated	3510	3512
# of MEROPS annotated	276	276
# of CAZymes annotated	388	389
# of Secondary metabolite clusters	72	71

revealed that *P. ucsense* 2HH possess a few orthologous of known fungal cell wall-associated proteins. The orthologous groups (Suppl. Table S04) also shows that there is a number of distinct cell wall-related proteins between the analysed fungi, allowing an investigation of conserved and specific elements between the analysed species.

Overall, cell wall-associated proteins remain conserved in the analysed fungi, especially those involved in biosynthesis of the main cell wall structural

components: chitin, β-1,3-glucan, lichenin and α-1,3-glucan. We also observed that the major regulatory proteins required for cell wall assembly and modifications are conserved. On the other hand, some components were found only in a specific analysed genus or a specific analysed species, particularly those involved in changing the cell wall during the fungus life cycle. Lastly, some cell wall components were found only in a specific analysed genus, such as some glycoproteins and proteins involved in biogenesis of the cell wall as a three-dimensional matrix.

When we compare these proteins of *P. ucsense* 2HH with *P. oxalicum* 114-2, the closest and available relative, the major differences in cell wall formation are components involved in changing the cell wall throughout the fungus life cycle. We observed differences in the number of coding genes for chitinase (GH18), β-acetylhexosaminidase (GH84), endo-1,3(4)-β-glucanase (GH16), glucan endo-1,3-β-D-glucosidase (GH16), glucan endo-1,6-β-glucosidase (GH5) and 1,3-β-glucosidase (GH55). These enzymes are required during the biogenesis of the cell wall, spore germination, hyphal branching and septum formation in filamentous fungi. Chitinases and β-hexosaminidases lead to re-modelling of the cell wall during growth and morphogenesis, while the other enzymes participate in the metabolism of β-D-glucan, the main structural component of the cell wall (Adams 2004).

P. ucsense 2HH presents five genes less of chitinases from the GH18 family than *P. oxalicum* 114-2. Differences in the number of proteins related to β-D-glucan metabolism are less expressive, but not less important. These changes in cell wall biosynthesis likely indicate modifications on growth and morphogenesis when compared to *P. oxalicum* 114-2. This is also consistent with observations of symbiotic co-adaptations between *Leucoagaricus gongylophorus* and leaf-cutting ants that led to increased chitin content and fungal cell wall thickness (Nygaard et al. 2016). In addition, we also observed an endo-1,6-β-glucosidase (GH5) and a β-acetylhexosaminidase (GH84) which apparently are pseudogenes, predicted by gene finders, but not conserved when aligned with other fungi.

Tracking mutations in the S1M29 mutant

The enzymes secreted by the 2HH wild-type provide an effective enzyme formulation for complete saccharification of plant residues, being rich in cellulases and hemicellulases (Schneider et al. 2016). The highest cellulases yields were produced in mutant S1M29 (Dillon et al. 2011; dos Reis et al. 2013; Schneider et al. 2016). Despite the efforts and advances achieved with classical mutagenesis, molecular knowledge of the mutations accumulated in the mutant S1M29 are important to design new strains. In this sense, ultra-deep WGS data of both strains allowed the accurate SNP calling, discovering variations between the mutant S1M29 and the parental strain 2HH. Both draft genomes were aligned using progressiveMauve resulting in 1337 SNPs, the most common type of genetic variation. We identified 8067 identical proteins in both strains, representing 98.7% of putative proteins. In addition, 52 proteins were affected by SNPs and contain amino-acid substitutions, representing 0.64% of the proteome. Alignment results and amino-acid substitutions are available in Suppl. Table S02.The origin tracking of the two main characteristics that differentiate the S1M29 mutant from the parental 2HH are presented below: (i) the hyperproduction of cellulases and (ii) the absence of green pigmentation in solid culture medium.

Origin tracking of the cellulase hyperproduction

There are no significant differences in the size of the genome and in the composition of proteins. The majority of the sequences of the CAZymes are also identical in the S1M29 mutant and in the 2HH wild-type. Also, we verified that the well-known transcription factors of *Penicillium* species, involved in cellulases and hemicellulases transcription regulation, have the same amino-acid sequences in both strains. Consequently, the variation on enzymatic production by both strains likely could be explained at the transcription level and not due to possible large-scale genomic rearrangements.

Table 3 Major proteins potentially involved in cellulase hyperproduction by the S1M29 mutant

Protein Id 2HH/S1M29	Mutation	Impact	Annotation
PECH_005648/PECM_002864	D194P	3	β-glucosidase BGL2
PECH_004634/PECM_006143	S259P	3	Developmental regulator FlbA
PECH_007011/PECM_006829	Y370D & I381E	2	Zn2Cys6 binuclear cluster domain
PECH_006810/PECM_005158	P277S	2	MFS-type transporter
PECH_002170/PECM_001690	L203F	2	Aromatic amino acid and leucine permease
PECH_008009/PECM_007813	S371F	2	Pre-mRNA-processing factor
PECH_006814/PECM_005162	A712D	2	Pre-mRNA-splicing helicase
PECH_008673/PECM_006103	Promoter region	2	MFS-type Sugar/inositol transporter

In order to classify the degree of the impact of the mutations in the improvement of cellulase expression of the mutant strain, we categorised proteins affected by SNPs into four levels of likely impact. The main proteins affected by the SNPs, classified at levels 3 and 2 of potential involvement on cellulase expression increment, are summarised in Table 3. In that, it is possible to verify the amino-acid substitution represented by the amino-acid in the parental 2HH, its location in the protein sequence and the respective mutation identified in the S1M29 mutant and the annotation associated.

Besides the mutations in BGL2 and FlbA, classified with a potential 3 of positive impact on the expression of cellulases, six other proteins were listed in Table 3, classified as level 2. This group comprises transporters and transcription factors that are commonly involved in numerous regulatory pathways, affecting directly or indirectly the cellular metabolism. In addition, pre-mRNA-splicing and pre-mRNA-processing factors may also play important roles in regulatory systems. While level 1 encompasses proteins related to RNA polymerase, RNA interference, gene silencing, and less influential transcription factors and transporters. Finally, level 0 covers proteins with no involvement in the cellulase expression system. Mutations classified as level 0 and 1 are shown in the Suppl. Table S02.

Overall, there were no significant changes in encoding CAZymes regions between the parental and the mutant. We found just four single amino-acid substitutions in CAZymes. The first three enzymes of glycoside hydrolase (GH) and glycosyltransferase (GT) families probably do not impact directly the cellulase expression: (i) Mannan endo-1,6-α-mannosidase (A5V); (ii) Glycoside Hydrolase Family 78 protein (A124T); and (iii) Glycosyl Transferase Family 90 protein (A287V).

The fourth CAZyme containing amino-acid substitution corresponds to a key finding of this study. A single amino-acid substitution occurs at position 194 in BGL2, changing an aspartic acid in the parental 2HH (PECH_005648) for a proline in the mutant S1M29 (PECM_002864). Considering this SNP of the major intracellular β-glucosidase BGL2, belonging to the GH1 family, it was reported that BGL2 orthologous play an important role in induction of cellulases in *Trichoderma reesei* and in *P. oxalicum* 114-2.

In a *T. reesei* mutant, it was observed that a single-nucleotide mutation of *bgl2* explains the enhanced cellulase expression, when the strains were cultivated on cellulose and cellobiose (Shida et al. 2015). Besides, in *P. oxalicum*, the major intracellular β-glucosidase BGL2 (PDE_00579) was found to be dependent on ClrB at the transcription level and *bgl2* deletion facilitates the synergistic expression of cellulase genes. Lack of this β-glucosidase facilitates the accumulation of intracellular cellodextrins, which can trigger signalling cascades that include expression of cellulase genes (Li et al. 2015; Yao et al. 2016).

As well as the mutation in *T. reesei*, in the S1M29 mutant the single amino-acid substitution is not in the catalytic domain of the enzyme BGL2. Surprisingly,

a single amino-acid substitution may be the major factor influencing the increase of cellulase expression in the S1M29 mutant. Our analysis led us to suggest that the SNP in *bgl2* may have reduced the activity of this enzyme, affording the hyperproduction of cellulases by the mutant S1M29. The effect of this amino-acid substitution on cellulase production could be also analysed by a combination of gene complementation and disruption techniques. However, these experiments are outside of the scope of this work.

Another key mutation is S259P identified in FlbA, which is a fungal Zn(II)2-Cys6 transcription regulator of conidiophore development. The conidiation is essential for industrial applications since the conidia are used as starters in the first step of fermentation. In this way, the regulatory pathway that controls conidiophore development and spore maturation is well-described. Various upstream developmental activators known as FLBs have been identified in filamentous fungi, reported to be involved in the regulation of conidiophore development (Van Munster et al. 2015). In *A. niger*, proteomics analysis indicates an increase in the number of total proteins and CAZymes during growth of the Δ*flbA* strain, suggesting that the *flbA* gene is an interesting target for improvement of industrial strains. Deletion of *flbA* in *A. nidulans* results in strongly reduced sporulation and excessive growth of aerial or submerged hyphae followed by autolytic collapse of the mycelium (Van Munster et al. 2015).

Experimental results from *P. oxalicum* (Li et al. 2015; Yao et al. 2016) and *T. reesei* (Shida et al. 2015) suggest that BGL2 likely is the major mutation involved in cellulase hyperproduction of the S1M29 mutant. However, experimental evidence of *A. niger* (Van Munster et al. 2015) suggests that the mutation in FlbA may also have an impact on the production of cellulases. Our results suggest that these mutations probably influence the regulatory pathways directly, resulting in overexpression of the main cellulases. However, other identified mutations may also have had a less positive influence on cellulase expression. Finally, we suggest that the group of mutations was responsible for the increased expression of cellulases in the mutant and not just one isolated mutation.

Origin tracking of albinism

The production of white instead of green pigmented conidia is a striking phenotypic character of the S1M29 mutant, motivating comparative analysis between the *P. ucsense* proteome and the proteins associated with the three melanin biosynthesis pathways, characterised in the following filamentous fungi: *A. fumigatus* Af293 (Teixeira et al. 2017), *A. niger* CBS 513.88 (Teixeira et al. 2017), *T. marneffei* ATCC 18224 (Woo et al. 2010), *P. chrysogenum* P2niaD18 (Woo et al. 2010).

Our comparative analyses revealed that *P. ucsense* possesses all the orthologous genes for the production of eumelanin by the DHN pathway, suggesting that the DHN pathway is the major route for melanin production in *P. ucsense*. In addition, we observed that L-tyrosine degradation is the most conserved pathway across the analysed fungi. In *A. niger* CBS 513.88, DOPA-melanin pathway is responsible for melanin production and this route is composed of tyrosinase coding genes and a vast number of laccase coding genes (Pal et al. 2014). In contrast to *A. niger*, we did not find homologues for these laccases in *P. ucsense* (Suppl. Table S03).

DHN-melanin genes are frequently encoded in biosynthetic gene clusters (Eisenman and Casadevall 2012). As presented in Fig. 3, we did not verify this organisation in *A. niger* CBS 513.88, *P. chrysogenum* P2niaD18, *P. oxalicum* 114-2 and *P. ucsense* 2HH. The DHN-melanin encoding genes were probably acquired by the common ancestor of *Penicillium* spp. and *Aspergillus* spp., and subsequent divergence and gene re-arrangement resulted in different gene orders and orientations of the individual genes in the different fungi (Woo et al. 2010).

In *A. fumigatus*, a cluster of six genes includes the *alb1* gene encoding a conidial pigment polyketide synthase, an initial precursor of DHN-melanin, involved in the production of the heptaketide naphtopyrone YWA1. Also in *A. fumigatus*, *alb1* disruption resulted in the albino conidial phenotype (Tsai et al. 1999). Analysing the DHN-melanin pathway of *P. ucsense* strains, we found the same amino-acid

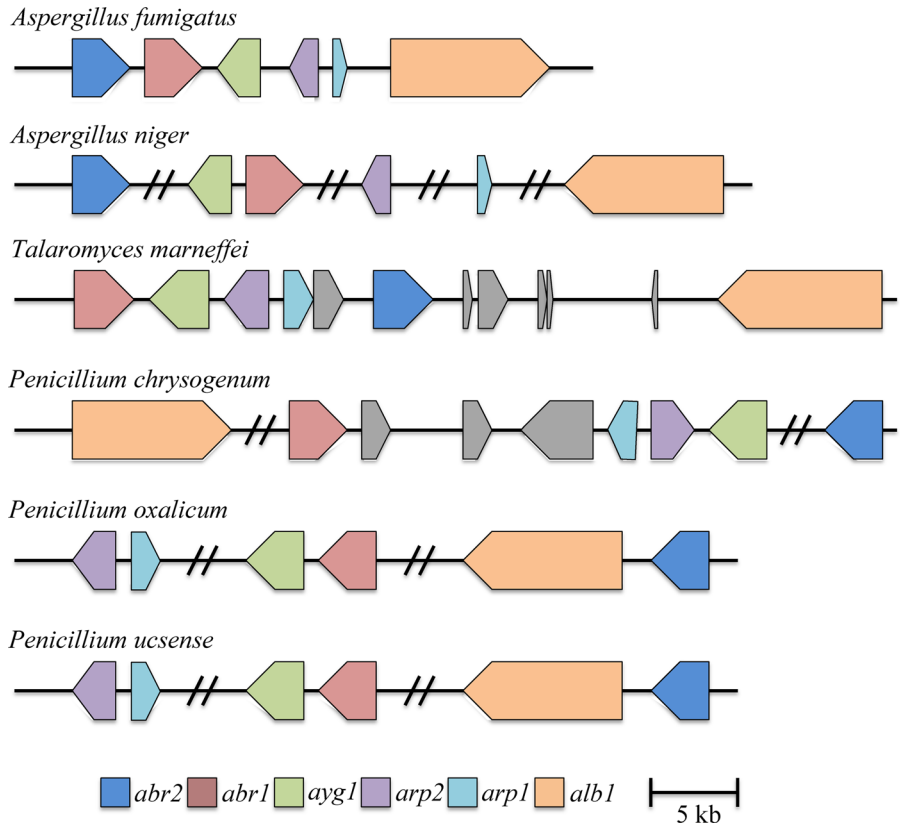


Fig. 3 Map of the DHN melanin-biosynthesis gene organisation in *Penicillium ucsense* 2HH and closely related fungi, based on Woo et al. (2010). *abr2*, conidial pigment biosynthesis oxidase; *abr1*, conidial pigment biosynthesis oxidase; *ayg1*, conidial pigment biosynthesis protein; *arp2*, conidial pigment biosynthesis 1,3,6,8-tetrahydroxynaphthalene reductase; *arp1*, conidial pigment biosynthesis scytalone dehydratase; *alb1*, conidial pigment polyketide synthase

sequences for all proteins, except a single amino-acid substitution at amino-acid 526 of ALB1, orthologue of *A. fumigatus* Af293. ALB1 exchanged glycine in the wild-type strain 2HH (PECH_008565) for aspartic acid in the mutant S1M29 (PECM_000136).

In order to observe the effect of this mutation, fungal morphology was monitored by macromorphological observations of both strains after 7 and 14 days of growth on four different growth media. In Fig. 4, we observe visible changes in conidium pigmentation in all growth media. While the 2HH wild-type produces characteristic green-coloured conidia, the conidia of the mutant are white (albino phenotype), similar to the *alb1* gene disruption observed in *A. fumigatus*. These results indicate that the DHN pathway is responsible for melanin production in *P. ucsense*.

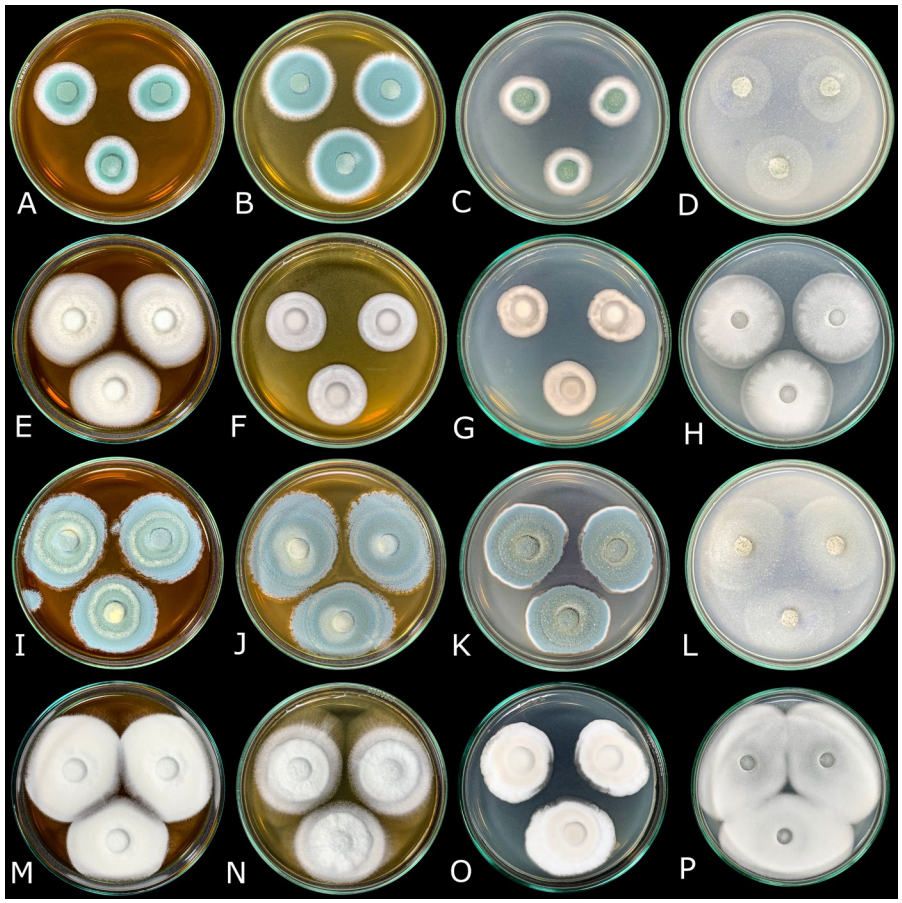


Fig. 4 Macromorphological observation photographs of the impact of the *alb1* mutation in *Penicillium ucsense* **a** 2HH colonies on MEA after 7 days at 28 °C. **b** 2HH colonies on MEAbl after 7 days at 28 °C. **c** 2HH colonies on CYA after 7 days at 28 °C. **d** 2HH colonies on IHMM after 7 days at 28 °C. **e** S1M29 colonies on MEA after 7 days at 28 °C. **f** S1M29 colonies on MEAbl after 7 days at 28 °C. **g** S1M29 colonies on CYA after 7 days at 28 °C. **h** S1M29 colonies on IHMM after 7 days at 28 °C. **i** 2HH colonies on MEA after 14 days at 28 °C. **j** 2HH colonies on MEAbl after 14 days at 28 °C. **k** 2HH colonies on CYA after 14 days at 28 °C. **l** 2HH colonies on IHMM after 14 days at 28 °C. **m** S1M29 colonies on MEA after 14 days at 28 °C. **n** S1M29 colonies on MEAbl after 14 days at 28 °C. **o** S1M29 colonies on CYA after 14 days at 28 °C. **p** S1M29 colonies on IHMM after 14 days at 28 °C

Conclusions

The description of *Penicillium* strain 2HH as a novel species in section *Lanata-Divaricata* is essential to carry out comparative studies with other fungal species. *Penicillium ucsense* 2HH belongs to series *Oxalica* and shares the production of high levels of cellulase with *P. oxalicum*, a species used for commercial production of cellulolytic enzymes in China. The six new strains of the same species collected from the marine environment in the Republic of Korea represent an important discovery related to the ecological preferences of this new species.

The genomes of *P. ucsense* 2HH/S1M29 have been deposited at GenBank. The genomic comparison of the mutant and the wild-type strains highlighted a wide set of mutations, of which only a few have been analysed in detail. BGL2 and FlbA likely are the major mutations involved in cellulase hyperproduction of the S1M29 mutant. Ongoing studies aim to identify the cellulolytic complex encoding genes and their main expression regulators,

comprising transcription factors and sugar transporters. Moreover, the single amino-acid substitution in ALB1, precursor enzyme of the DHN-melanin biosynthesis, is the responsible for the loss of pigmentation of the conidia, evidencing that DHN-melanin biosynthesis pathway is the major responsible for melanin production in *P. ucsense*. The composition of cell wall-associated proteins of *P. ucsense* 2HH shows considerable differences in the number of proteins, when compared to *P. oxalicum* 114-2, the most closely related genome sequenced species. The major differences comprise less chitinases, proteins related to β-D-glucan metabolism and two potential pseudogenes.

Finally, our study provides an important step in building toward understanding species distribution, molecular machinery, cellulolytic system, melanin production and cell wall biosynthesis of this new *Penicillium* species.

Acknowledgements We would like to thank the Coordination of Improvement of Higher Education Personnel (CAPES), National Council for Scientific and Technological Development (CNPq), the Bahia State University (UNEB) and the University of Caxias do Sul (UCS).

Author contributions Conceptualization, methodology, investigation, visualisation, writing—original draft, review and editing: [ARL]; Assembly and annotation of genomes: [ARL, NSO and FPA]; Software for genome annotation and deposit of genomes: [EB]; Macro- and micromorphological characterization: [JH]; Collection and sequencing of molecular markers from strains of the Republic of Korea [MSP and YWL]; Macromorphological observation of albino phenotype: [RCF]; Supervision: [SAS]; Project administration, funding acquisition and resources: [MC and AJPD].

Funding We are grateful to the Coordination for the Improvement of Higher Education Personnel (CAPES) for the DSc scholarship (88887.158496/2017-00 to ARL). This research was supported by grants from CAPES (3255/2013) and the National Council for Scientific and Technological Development (CNPq) (472153/2013-7). MC and AJPD are CNPq Research Fellowship. We are grateful to Bahia State University (UNEB) for the leave of absence (3.145/2016 to ARL) and financial support.

Data Availability Data supporting the results reported in the article can be found in Supplementary Data or publicly available at NCBI databases.

Declarations

Conflict of interest All authors declare no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

Adams DJ (2004) Fungal cell wall chitinases and glucanases. Microbiology+ 150:2029–35. https://doi.org/10.1099/mic.0.26980-0

Aguilar-Pontes MV, Brandl J, McDonnell E et al (2018) The gold-standard genome of *Aspergillus niger* NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. Stud Mycol 91:61–78. https://doi.org/10.1016/j.simyco.2018.10.001

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z

Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S (2015) FastQC v0.11.5. Computer program and documentation distributed by Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA-UK 6:11. https://doi.org/10.1186/s13100-015-0041-9

Bernard M, Latgé JP (2001) *Aspergillus fumigatus* cell wall: composition and biosynthesis. Med Mycol 39:9–17. https://doi.org/10.1080/mmy.39.1.9.17

Blin K, Medema MH, Kottmann R, Lee SY, Weber T (2017) The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 45:D555–D559. https://doi.org/10.1093/nar/gkw960

Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421

Camassola M, Dillon AJP (2006) Effect of methylxanthines on production of cellulases by *Penicillium echinulatum*. Appl Microbiol 102:478–485. https://doi.org/10.1111/j.1365-2672.2006.03098.x

Camassola M, Dillon AJP (2007) Production of cellulases and hemicellulases by *Penicillium echinulatum* grown on pretreated sugar cane bagasse and wheat bran in solid-state fermentation. Appl Microbiol 103:2196–2204. https://doi.org/10.1111/j.1365-2672.2007.03458.x

Camassola M, Dillon AJP (2009) Biological pretreatment of sugar cane bagasse for the production of cellulases and xylanases by *Penicillium echinulatum*. Ind Crop Prod 29:642–647. https://doi.org/10.1016/j.indcrop.2008.09.008

Camassola M, Dillon AJP (2010) Cellulases and xylanases production by *Penicillium echinulatum* grown on sugar cane bagasse in solid-state fermentation. Appl

Biochem Biotech 162:1889–1900. https://doi.org/10.1007/s12010-010-8967-3

Camassola M, Dillon AJP (2012) Steam-exploded sugar cane bagasse for on-site production of cellulases and xylanases by *Penicillium echinulatum*. Energ Fuel 26:5316–5320. https://doi.org/10.1021/ef3009162

Camassola M, De Bittencourt LR, Shenem NT, Andreaus J, Dillon AJP (2004) Characterization of the cellulase complex of *Penicillium echinulatum*. Biocatal Biotransfor 22:391–396. https://doi.org/10.1080/10242420400024532

Carrau JL, Dillon AJP, Ribeiro RTS, Leygue-Alba NMR, Azevedo JL (1981) Produção de enzimas celulolíticas por microrganismos. In: Simpósio Internacional de Engenharia Genética. Piracicaba, SP. pp 39.

Chikhi R, Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31–37. https://doi.org/10.1093/bioinformatics/btt310

Chomczynski P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques 15:532–537

Darling AE, Mau B, Perna NT (2010) Progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5:e11147. https://doi.org/10.1371/journal.pone.0011147

Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: More models, new heuristics and parallel computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109

Dillon AJP, Zorgi C, Camassola M, Henriques JAP (2006) Use of 2-deoxyglucose in liquid media for the selection of mutant strains of *Penicillium echinulatum* producing increased cellulase and β-glucosidase activities. Appl Microbiol Biot 70:740–746. https://doi.org/10.1007/s00253-005-0122-7

Dillon AJP, Bettio M, Pozzan FG, Andrighetti T, Camassola M (2011) A new *Penicillium echinulatum* strain with faster cellulase secretion obtained using hydrogen peroxide mutagenesis and screening with 2-deoxyglucose. Appl Microbiol 111:48–53. https://doi.org/10.1111/j.1365-2672.2011.05026.x

dos Reis L, Fontana RC, da Silva DP et al (2013) Increased production of cellulases and xylanases by *Penicillium echinulatum* S1M29 in batch and fed-batch culture. Bioresource Technol 146:597–603. https://doi.org/10.1016/j.biortech.2013.07.124

Eisenman HC, Casadevall A (2012) Synthesis and assembly of fungal melanin. Appl Microbiol Biot 93:931–940. https://doi.org/10.1007/s00253-011-3777-2

El-Gebali S, Mistry J, Bateman A et al (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995

Epstein SC, Charkoudian LK, Medema MH (2018) A standardized workflow for submitting data to the Minimum Information about a Biosynthetic Gene cluster (MIBiG) repository: prospects for research-based educational experiences. Stand Genomic Sci 13:16. https://doi.org/10.1186/s40793-018-0318-y

Free SJ (2013) Fungal cell wall organization and biosynthesis. In: Friedmann T, Dunlap JC, Goodwin SF (ed) Advances in genetics. Academic Press, New York, pp 33–82. https://doi.org/10.1016/B978-0-12-407677-8.00002-6

Gostinčar C (2020) Towards genomic criteria for delineating fungal species. J Fungi (Basel) 6:246. https://doi.org/10.3390/jof6040246

Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Green MR, Sambrook JF (2012) Molecular cloning: a laboratory manual, Vol. 1. 4th ed. Press CSHL, Cold Springs Harbour Press, New York.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haas BJ, Salzberg SL, Zhu W et al (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7

Houbraken J, Kocsubé S, Visagie CM et al (2020) Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (*Eurotiales*): an overview of families, genera, subgenera, sections, series and species. Stud Mycol 95:5–169. https://doi.org/10.1016/j.simyco.2020.05.002

Huerta-Cepas J, Forslund K, Coelho LP et al (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148

Huerta-Cepas J, Szklarczyk D, Heller D et al (2019) EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085

Hyde KD, Xu J, Rapior S et al (2019) The amazing potential of fungi: 50 ways we can exploit fungi industrially. Fungal Divers 97:1–136. https://doi.org/10.1007/s13225-019-00430-9

Inkscape (2019) v. 0.92.2 Computer program and documentation distributed by the authors. https://inkscape.org

Johnson LS, Eddy SR, Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431. https://doi.org/10.1186/1471-2105-11-431

Jones P, Binns D, Chang HY et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070

Katoh K, Rozewicki J, Yamada KD (2018) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Kent WJ (2002) BLAT–-the BLAST-like alignment tool. Genome Res 12:656–664. https://doi.org/10.1101/gr.229202

Kim D, Langmead B, Salzberg SL (2015) HISAT: afast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317

Kriventseva EV, Tegenfeldt F, Petty TJ et al (2015) OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res 43:D250–D256. https://doi.org/10.1093/nar/gku1220

Krueger F (2015) Trim Galore! v0.4.4. Computer program and documentation distributed by Babraham Institute. https://github.com/FelixKrueger/TrimGalore

Kubátová A, Hujslová M, Frisvad JC, Chudíčková M, Kolařík M (2019) Taxonomic revision of the biotechnologically important species *Penicillium oxalicum* with the description of two new species from acidic and saline soils. Mycol Prog 18:215–228. https://doi.org/10.1007/s11557-018-1420-7

Langfelder K, Streibel M, Jahn B, Haase G, Brakhage AA (2003) Biosynthesis of fungal melanins and their importance for human pathogenic fungi. Fungal Genet Biol 38:143–158. https://doi.org/10.1016/S1087-1845(02)00526-1

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lechner M, Findeiß S, Steiner L et al (2011) Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124. https://doi.org/10.1186/1471-2105-12-124

Lenz AR, Galán-Vasquez E, Balbinot E et al (2020) Gene regulatory networks of *Penicillium echinulatum* 2HH and *Penicillium oxalicum* 114–2 inferred by a computational biology approach. Front Microbiol 11:588263. https://doi.org/10.3389/fmicb.2020.588263

Lenz AR, Balbinot E, Oliveira NS et al (2022) Analysis of carbohydrate-active enzymes and sugar transporters in *Penicillium echinulatum*: a genome-wide comparative study of the fungal lignocellulolytic system. Gene 822:146345. https://doi.org/10.1016/j.gene.2022.146345

Li Z, Yao G, Wu R et al (2015) Synergistic and dose-controlled regulation of cellulase gene expression in *Penicillium oxalicum*. PLoS Genet 11:e1005509–e1005509. https://doi.org/10.1371/journal.pgen.1005509

Liu G, Zhang L, Wei X et al (2013) Genomic and secretomic analyses reveal unique features of the lignocellulolytic enzyme system of *Penicillium decumbens*. PLoS ONE 8:e55185. https://doi.org/10.1371/journal.pone.0055185

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413

McDonnell E, Strasser K, Tsang A (2018) Manual gene curation and functional annotation. In: de Vries RP, Tsang A, Grigoriev IV (ed) Methods in molecular biology. New York, NY, pp 185–208. https://doi.org/10.1007/978-1-4939-7804-5_16

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: IEEE Gateway Computing Environments workshop (GCE), pp 1–8. https://doi.org/10.1109/GCE.2010.5676129

Moller S, Croning MDR, Apweiler R (2002) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 18:218–218. https://doi.org/10.1093/bioinformatics/18.1.218

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185. https://doi.org/10.1093/nar/gkm321

Novello M, Vilasboa J, Schneider WDH, dos Reis L, Fontana RC, Camassola M (2014) Enzymes for second generation ethanol: Exploring new strategies for the use of xylose. RSC Adv 4:21361–21368. https://doi.org/10.1039/c4ra00909f

Nygaard S, Hu H, Li C et al (2016) Reciprocal genomic evolution in the ant-fungus agricultural symbiosis. Nat Commun 7:12233. https://doi.org/10.1038/ncomms12233

Pal AK, Gajjar DU, Vasavada AR (2014) DOPA and DHN pathway orchestrate melanin synthesis in *Aspergillus* species. Med Mycol 52:10–18. https://doi.org/10.3109/13693786.2013.826879

Park MS, Oh SY, Lee S, Eimes JA, Lim YW (2018) Fungal diversity and enzyme activity associated with sailfin sandfish egg masses in Korea. Fungal Ecol 34:1–9. https://doi.org/10.1016/j.funeco.2018.03.004

Parkin EA (1940) The digestive enzymes of some wood-boring beetle larvae. Exp Biol 17:364–377. https://doi.org/10.1242/jeb.17.4.364

Patel PK, Free SJ (2019) The genetics and biochemistry of cell wall structure and synthesis in *Neurospora crassa*, a model filamentous fungus. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.02294

Peng M, Aguilar-Pontes MV, de Vries RP, Mäkelä MR (2018) In silico analysis of putative sugar transporter genes in *Aspergillus niger* using phylogeny and comparative transcriptomics. Front Microbiol 9:01045. https://doi.org/10.3389/fmicb.2018.01045

Pralea IE, Moldovan RC, Petrache AM et al (2019) From extraction to advanced analytical methods: the challenges of melanin analysis. Int J Mol Sci 20:3943. https://doi.org/10.3390/ijms20163943

Rambaut A (2009) FigTree v.1.4.4. Computer program and documentation distributed by the author. http://tree.bio.ed.ac.uk/software/

Ribeiro DA, Cota J, Alvarez TM et al (2012) The *Penicillium echinulatum* Secretome on Sugar Cane Bagasse. PLoS ONE 7:e50571–e50571. https://doi.org/10.1371/journal.pone.0050571

Ronquist F, Teslenko M, van der Mark P et al (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Rubini MR, Dillon AJP, Kyaw CM, Faria FP, Poças-Fonseca MJ, Silva-Pereira I (2010) Cloning, characterization and heterologous expression of the first *Penicillium echinulatum* cellulase gene. Appl Microbiol 108:1187–1198. https://doi.org/10.1111/j.1365-2672.2009.04528.x

Schneider WDH, dos Reis L, Camassola M, Dillon AJP (2014) Morphogenesis and production of enzymes by *Penicillium echinulatum* in response to different carbon sources.

Biomed Res Int 2014:10. https://doi.org/10.1155/2014/254863
Schneider WDH, Gonçalves TA, Uchima CA et al (2016) *Penicillium echinulatum* secretome analysis reveals the fungi potential for degradation of lignocellulosic biomass. Biotechnol Biofuels 9:66. https://doi.org/10.1186/s13068-016-0476-3
Schneider WDH, Gonçalves TA, Uchima CA et al (2018) Comparison of the production of enzymes to cell wall hydrolysis using different carbon sources by *Penicillium echinulatum* strains and its hydrolysis potential for lignocelullosic biomass. Process Biochem 66:162–170. https://doi.org/10.1016/j.procbio.2017.11.004
Schneider WDH, Fontana RC, Baudel HM et al (2020) Lignin degradation and detoxification of eucalyptus wastes by on-site manufacturing fungal enzymes to enhance second-generation ethanol yield. Appl Energ 262:114493. https://doi.org/10.1016/j.apenergy.2020.114493
Shida Y, Yamaguchi K, Nitta M et al (2015) The impact of a single-nucleotide mutation of *bgl2* on cellulase induction in a *Trichoderma reesei* mutant. Biotechnol Biofuels 8:230. https://doi.org/10.1186/s13068-015-0420-y
Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31. https://doi.org/10.1186/1471-2105-6-31
Smit A, Hubley R, Green P (2013) RepeatMasker Open-4.0. Computer program and documentation distributed by the authors. http://www.repeatmasker.org
Solovyev V (2008) Statistical Approaches in Eukaryotic Gene Prediction. In: Balding DJ, Bishop M and Cannings C (ed). Handbook of Statistical Genetics. Third edition. Wiley Online Library, Chapter 4, pp 97–159. https://doi.org/10.1002/9780470061619.ch4
Stamatakis A (2014) RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62
Teixeira MM, Moreno LF, Stielow BJ et al (2017) Exploring the genomic diversity of black yeasts and relatives (*Chaetothyriales*, *Ascomycota*). Stud Mycol 86:1–28. https://doi.org/10.1016/j.simyco.2017.01.001
Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18:1979–1990. https://doi.org/10.1101/gr.081612.108
The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049
The GIMP (2020) GIMP v2.10.14. Computer program and documentation distributed by Development Team. https://www.gimp.org
Tsai HF, Wheeler MH, Chang YC, Kwon-Chung KJ (1999) A developmentally regulated gene cluster involved in conidial pigment biosynthesis in *Aspergillus fumigatus*. J Bacteriol 181:6469–6477. https://doi.org/10.1128/jb.181.20.6469-6477.1999
Van Munster JM, Nitsche BM, Akeroyd M et al (2015) Systems approaches to predict the functions of glycoside hydrolases during the life cycle of *Aspergillus niger* using developmental mutants Δ*brlA* and Δ*flbA*. PLoS ONE 10:e0116269. https://doi.org/10.1371/journal.pone.0116269
Villesen P (2007) FaBox: an online toolbox for FASTA sequences. Mol Ecol Notes 7:965–968. https://doi.org/10.1111/j.1471-8286.2007.01821.x
Visagie CM, Houbraken J, Frisvad JC et al (2014) Identification and nomenclature of the genus *Penicillium*. Stud Mycol 78:343–371. https://doi.org/10.1016/j.simyco.2014.09.001
Waterhouse RM, Seppey M, Simao FA et al (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. https://doi.org/10.1093/molbev/msx319
Weirauch MT, Yang A, Albu M et al (2014) Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158:1431–1443. https://doi.org/10.1016/j.cell.2014.08.009
Woo PCY, Tam EWT, Chong KTK et al (2010) High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. FEBS J 277:3750–3758. https://doi.org/10.1111/j.1742-4658.2010.07776.x
Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ (2016) GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. In: Mathé E, Davis S (ed) Methods in molecular biology. New York, NY, pp 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15
Yao G, Wu R, Kan Q et al (2016) Production of a high-efficiency cellulase complex via β-glucosidase engineering in *Penicillium oxalicum*. Biotechnol Biofuels 9:78. https://doi.org/10.1186/s13068-016-0491-4
Zhang H, Yohe T, Huang L et al (2018) DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95-101. https://doi.org/10.1093/nar/gky418