genomeA
Journals.ASM.org

Draft Genome Sequence of the Fungus *Penicillium brasilianum* MG11

Fabian Horn,[a*] Jörg Linde,[a] Derek J. Mattern,[b] Grit Walther,[c] Reinhard Guthke,[a] Axel A. Brakhage,[b,e] Vito Valiante[b,d]

Departments of Systems Biology/Bioinformatics[a] and Molecular and Applied Microbiology,[b] National Center for Invasive Mycoses,[c] and Leibniz Junior Research Group—Biobricks of Microbial Natural Product Syntheses,[d] Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (HKI), Jena, Germany; Friedrich Schiller University, Institute for Microbiology, Jena, Germany[e]

* Present address: Fabian Horn, Section 4.5 Geomicrobiology, GFZ German Centre For GeoSciences, Potsdam, Germany.

The genus *Penicillium* belongs to the phylum *Ascomycota* and includes a variety of fungal species important for food and drug production. We report the draft genome sequence of *Penicillium brasilianum* MG11. This strain was isolated from soil, and it was reported to produce different secondary metabolites.

Received 26 May 2015 **Accepted** 24 July 2015 **Published** 3 September 2015

Citation Horn F, Linde J, Mattern DJ, Walther G, Guthke R, Brakhage AA, Valiante V. 2015. Draft genome sequence of the fungus *Penicillium brasilianum* MG11. Genome Announc 3(5):e00724-15. doi:10.1128/genomeA.00724-15.

Address correspondence to Vito Valiante, vito.valiante@leibniz-hki.de.

Penicillium is one of the most used genera in biotechnology. It contains well-known species such as *Penicillium chrysogenum*, the major producer of penicillin, and *Penicillium camemberti* and *Penicillium roqueforti*, used in cheese manufacturing (1). *Penicillium brasilianum* is a fungus that grows in soil (2) and on decaying plant material (3, 4). The fungus has been previously reported to produce a wide range of secondary metabolites (5), which show antiviral, cytotoxic, carcinogenic, and insecticidal bioactivity (6, 7). Moreover, the fungus produces a number of different degrading enzymes of biotechnological interest (8).

We sequenced the *P. brasilianum* MG11 strain (7). Phylogenetic analysis of β-tubulin gene sequences, including sequences of ex-type or reference strains of *P. brasilianum*, confirmed the classification of this fungus.

Three different DNA libraries (paired end [PE], 2-kbp mate pair [MP], and 8-kbp MP) of *P. brasilianum* MG11 were sequenced using 2 × 300-bp Illumina MiSeq V3. Reads were adapter clipped, quality trimmed, error corrected (9), and digitally normalized (10). The genome was assembled using the AllPaths-LG assembler (11) and refined using SOAP GapCloser (12) and SEQuel (13). For RNA sequencing, total RNA was extracted from cultures grown in minimal medium for 18 h (shaking cultures) and 72 h (standing cultures) at 30°C. For each sample three biological replicates were sequenced using 2 × 100-bp Illumina HiSeq 2000 (LGC Genomics, Berlin, Germany).

Structural and functional gene prediction followed the protocol previously described (14–16). For the phylogeny-based prediction, the proteomes of *P. bilaiae*, *P. brevicompactum*, *P. chrysogenum*, *P. digitatum*, *P. glabrum*, *P. janthinellum*, *P. lanosocoeruleum*, and *P. raistrickii* (all available at the Joint Genome Institute [JGI]) were mapped to the reference genome. In contrast to other reported studies (15, 16), rRNA was removed from the RNA-Seq data using riboPicker (17). Functional annotations were obtained by using Blast2GO (18) and InterProScan (19) and included the prediction of secondary metabolite gene clusters using SMURF (20). Functional names were suggested by using BLAST and the fungal UniProt Knowledgebase (21).

DNA sequencing of all libraries resulted in 202.6 million reads (15.2 Gbp), from which 130.8 million reads (7.2 Gbp; estimated 205-fold genome coverage) were used for the assembly. RNA sequencing of all libraries resulted in 278.5 million reads (28.0 Gbp), from which 188.2 million reads (18.9 Gbp; estimated 526-fold genome coverage) passed quality filtering and were used for annotation. The assembly procedure resulted in a genome comprising 87 scaffolds (35.9 Mbp; N_{50}, 3.4 Mbp; N_{90}, 632.3 kbp). The genome annotation predicted 11,432 genes and 12,343 transcripts. The coding density of the genome is approximately 48%. InterPro domains were assigned to 9,748 transcripts, whereof 3,033 transcripts encode proteins with a predicted transmembrane domain. In 6,173 transcripts at least one gene ontology (GO) annotation was found. GO annotation was made accessible at FungiFun2 (22). Functional annotation suggested names for 4,758 transcripts and assigned 438 transcripts to 36 putative secondary metabolite gene clusters. The annotation contained 477 core eukaryotic genes (23).

Nucleotide sequence accession numbers. This genome project was uploaded to DDBJ/ENA/GenBank and is available under the accession numbers CDHK01000001 to CDHK01000087. This paper describes the first version of the genome. Genome data and additional information are also available at the HKI Genome Resource (http://www.genome-resource.de/).

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) CRC/Transregio 124 "Pathogenic fungi and their human host: networks of interaction," subproject INF (J.L.), the DFG-funded excellence graduate school Jena School for Microbial Communication (JSMC) (D.J.M., A.A.B.), and the Leibniz Research Cluster Biotechnology 2020$^+$ (V.V.).

We are grateful to Hideo Hayashi for providing us with the strain used in this work.

REFERENCES

1. **Le Dréan G, Mounier J, Vasseur V, Arzur D, Habrylo O, Barbier G.** 2010. Quantification of *Penicillium camemberti* and *P. roqueforti* mycelium by real-

time PCR to assess their growth dynamics during ripening cheese. Int J Food Microbiol **138:**100–107. doi:http://dx.doi.org/10.1016/ j.ijfoodmicro.2009.12.013.

2. **Christensen M, Frisvad JC, Tuthill DE**. 2000. *Penicillium* species diversity in soil and some taxonomic and ecological notes, p 309–320. *In* Samson RA, Pitt JI (ed), *Integration of modern taxonomic methods for Penicillium and Aspergillus classification*. Harwood Academic Publishers, Amsterdam, The Netherlands.
3. **Sang MK, Han GD, Oh JY, Chun S, Kim KD**. 2014. *Penicillium brasilianum* as a novel pathogen of onion (*Allium cepa L.*) and other fungi predominant on market onion in Korea. Crop Protect **65:**138–142. http://dx.doi.org/10.1016/j.cropro.2014.07.016.
4. **Cho HS, Hong SB, Go SJ**. 2005. First report of *Penicillium brasilianum* and *P. daleae* isolated from soil in Korea. Mycobiology **33:**113–117. http://dx.doi.org/10.4489/MYCO.2005.33.2.113.
5. **Tang HY, Zhang Q, Li H, Gao JM**. 2015. Antimicrobial and allelopathic metabolites produced by *Penicillium brasilianum*. Nat Prod Res **29:**345–348. http://dx.doi.org/10.1080/14786419.2014.940347.
6. **Schürmann BTM, Sallum WST, Takahashi JA**. 2010. Austin, dehydroaustin and other metabolites from *Penicillium brasilianum*. Química Nova **33**:1044–1046. http://dx.doi.org/10.1590/S0100-40422010000500007.
7. **Kataoka S, Furutani S, Hirata K, Hayashi H, Matsuda K**. 2011. Three Austin family compounds from *Penicillium brasilianum* exhibit selective blocking action on cockroach nicotinic acetylcholine receptors. Neurotoxicology **32:**123–129. http://dx.doi.org/10.1016/j.neuro.2010.10.003.
8. **Panagiotou G, Olavarria R, Olsson L**. 2007. *Penicillium brasilianum* as an enzyme factory; the essential role of feruloyl esterases for the hydrolysis of the plant cell wall. J Biotechnol **130:**219–228. http://dx.doi.org/10.1016/j.jbiotec.2007.04.011.
9. **Liu Y, Schröder J, Schmidt B**. 2013. Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics **29:**308–315. http://dx.doi.org/10.1093/bioinformatics/bts690.
10. **Brown CT, Crusoe MR, Edvenson G, Fish J, Adina H, McDonald E, Nahum J, Nanlohy K, Ortiz-Zuazaga H, Pell Jason SJ, Scott C, Srinivasan RR, Zhang Q**. 2014. The Khmer software package: enabling efficient sequence analysis. http://dx.doi.org/10.6084/m9.figshare.979190.
11. **Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB**. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A **108:**1513–1518. http://dx.doi.org/10.1073/pnas.1017351108.
12. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J**. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1:**18. http://dx.doi.org/10.1186/2047-217X-1-18.
13. **Ronen R, Boucher C, Chitsaz H, Pevzner P**. 2012. SEQuel: improving the accuracy of genome assemblies. Bioinformatics **28:**i188–i196. http://dx.doi.org/10.1093/bioinformatics/bts219.
14. **Linde J, Schwartze V, Binder U, Lass-Flörl C, Voigt K, Horn F**. 2014. *De novo* whole-genome sequence and genome annotation of *Lichtheimia ramosa*. Genome Announc **2**(5):e00888-14. http://dx.doi.org/10.1128/genomeA.00888-14.
15. **Horn F, Üzüm Z, Möbius N, Guthke R, Linde J, Hertweck C**. 2014. Draft genome sequences of symbiotic and nonsymbiotic *Rhizopus microsporus* strains CBS 344.29 and ATCC 62417. Genome Announc **3**(1):e01370-14. http://dx.doi.org/10.1128/genomeA.01370-14.
16. **Linde J, Duggan S, Weber M, Horn F, Sieber P, Hellwig D, Riege K, Marz M, Martin R, Guthke R, Kurzai O**. 2015. Defining the transcriptomic landscape of *Candida glabrata* by RNA-Seq. Nucleic Acids Res **43:**1392–1406. http://dx.doi.org/10.1093/nar/gku1357.
17. **Schmieder R, Lim YW, Edwards R**. 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics **28:**433–435. http://dx.doi.org/10.1093/bioinformatics/btr669.
18. **Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:**3674–3676. http://dx.doi.org/10.1093/bioinformatics/bti610.
19. **Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R**. 2005. InterProScan: protein domains identifier. Nucleic Acids Res **33:**W116–W120. http://dx.doi.org/10.1093/nar/gki442.
20. **Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND**. 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol **47:**736–741. http://dx.doi.org/10.1016/j.fgb.2010.06.003.
21. **The UniProt Consortium**. 2014. Activities at the universal protein resource (UniProt). Nucleic Acids Res **42:**D191–D198. http://dx.doi.org/10.1093/nar/gkt1140.
22. **Priebe S, Kreisel C, Horn F, Guthke R, Linde J**. 2015. FungiFun2: a comprehensive online resource for systematic analysis of gene lists from fungal species. Bioinformatics **31:**445–446. http://dx.doi.org/10.1093/bioinformatics/btu627.
23. **Parra G, Bradnam K, Korf I**. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23:**1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.