

Draft Genome Sequences of Fungus *Aspergillus calidoustus*

Fabian Horn,^{a,*} Jörg Linde,^a Derek J. Mattern,^{b,c} Grit Walther,^d Reinhard Guthke,^a Kirstin Scherlach,^e Karin Martin,^f Axel A. Brakhage,^{b,c} Lutz Petzke,^g Vito Valiante^h

Department of Systems Biology/Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (HKI), Jena, Germany^a; Department of Molecular and Applied Microbiology, HKI, Jena, Germany^b; Institute for Microbiology, Friedrich Schiller University, Jena, Germany^c; National Reference Center for Invasive Fungal Infections, HKI, Jena, Germany^d; Department of Biomolecular Chemistry, HKI, Jena, Germany^e; Bio Pilot Plant, HKI, Jena, Germany^f; BASF SE, Ludwigshafen, Germany^g; Leibniz Junior Research Group-BioBricks of Microbial Natural Product Syntheses, HKI, Jena, Germany^h

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

Here, we report the draft genome sequence of *Aspergillus calidoustus* (strain SF006504). The functional annotation of *A. calidoustus* predicts a relatively large number of secondary metabolite gene clusters. The presented genome sequence builds the basis for further genome mining.

Received 25 January 2016 Accepted 28 January 2016 Published 10 March 2016

Citation Horn F, Linde J, Mattern DJ, Walther G, Guthke R, Scherlach K, Martin K, Brakhage AA, Petzke L, Valiante V. 2016. Draft genome sequences of fungus *Aspergillus calidoustus*. *Genome Announc* 4(2):e00102-16. doi:10.1128/genomeA.00102-16.

Copyright © 2016 Horn et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

Aspergillus is the most studied filamentous genus in the ascomycetes division. It contains well-known human pathogens (e.g., *A. fumigatus*, *A. terreus*), fermentation agents of Asian food (e.g., *A. oryzae*), and different industrial producers (e.g., *A. niger* and *A. flavus*) (1). Additionally, the genus is well studied because of its immense source of natural products (2). The worldwide distributed *Aspergillus calidoustus* was recently separated from the mesophilic species *A. ustus* (3). *A. calidoustus* was predominantly isolated from indoor environments and immunocompromised patients (3, 4).

Strain SF006504 was identified as *A. calidoustus* based on the β -tubulin sequence (100% similarity with the ex-type strain CBS 121601). Genomic DNA was obtained from a sample cultured in *Aspergillus* minimal medium (AMM) (5). For Illumina MiSeq V2 DNA sequencing, four different libraries were prepared (paired-end [PE], 2kbp-MP, 5kbp-MP, 8kbp-MP). The genome assembly was generated using Allpaths-LG (6). RNA-seq data was obtained for three replicates for standard growth, hypoxic, and iron depletion conditions. Sequencing was performed with Illumina HiSeq 2000 (LGC Genomics, Berlin).

For gene prediction, we customized and augmented the pipeline of Haas et al. (7) as previously described (8), applying *ab initio* prediction tools, tools incorporating transcriptome data, and protein alignments. If applicable, the parameter sets for each tool were trained using the transcriptome assembly. All gene predictions were combined using EvidenceModeller and PASA (9).

For *ab initio* gene prediction, we applied GeneMark-ES (10), Augustus (11), and SNAP (12). Transcriptome data was incorporated into Augustus (11), FGENESH (13), and PASA (14), which utilized genome-guided and genome-independent transcriptome assemblies produced by Trinity (15) and Cufflinks (16).

Protein alignments were obtained by mapping protein sequences from *A. fumigatus*, *A. nidulans*, *A. oryzae*, *A. terreus*, and *A. niger* (downloaded from the *Aspergillus* Genome Database [17]), using Exonerate (18) and Scipio (19).

Functional annotation was performed using Blast2GO (20) and InterproScan (21). Gene descriptions were obtained by blasting the predicted protein sequences against fungal UniProt KnowledgeBase. Matches with e-values below 10^{-5} , 70% sequence identity, and a subject hit length of 70% were considered as highly similar. Secondary metabolite gene clusters were predicted using SMURF (22).

DNA-sequencing resulted in 59,066,664 raw reads, where 50,376,036 reads passed our quality-filter (estimated genome coverage, 300-fold) and have been used for genome assembly. The resulting assembly consists of 78 scaffolds and 41.1 Mbp (N_{50} 3.2Mbp; N_{90} 493 kbp). The total G+C content was 51%. RNA-sequencing resulted in a total of 393,543,839 raw reads and 352,698,587 preprocessed reads (estimated genome coverage, 850-fold). The final structural gene prediction resulted in 15,139 gene models and 15,537 transcripts. 484 eukaryotic core proteins were identified using CEGMA (23). The coding density of the genome was 60%. We assigned functional names to 5,572 transcripts, gene ontology (GO) categories to 8,352 transcripts, and protein domains to 13,610 transcripts. 3,771 transcripts were predicted to contain transmembrane domains, and 749 transcripts have been assigned to 53 secondary metabolite gene clusters. GO annotations have been made available for downstream analysis at FungiFun2 (24).

Nucleotide sequence accession numbers. This genome project was uploaded to DDBJ/ENA/GenBank and is available under accession numbers [CDMC01000001](https://www.ncbi.nlm.nih.gov/nuccore/CDMC01000001) to [CDMC01000078](https://www.ncbi.nlm.nih.gov/nuccore/CDMC01000078). 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

We and our associated institutions declare a financial interest. We applied for a patent related to the presented material and data.

FUNDING INFORMATION

This work was supported by BASF SE (Ludwigshafen) and the Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany. J.L. work was supported by the Deutsche Forschungsgemeinschaft (DFG) CRC/Transregio 124 "Pathogenic fungi and their human host: networks of interaction," project INF. D.J.M. was supported by the DFG-funded excellence graduate school Jena School for Microbial Communication (JSMC).

REFERENCES

- Bennett JW. 2010. An overview of the genus *Aspergillus*. In *Aspergillus: molecular biology and genomics*. Caister Academic Press, Portland.
- Brakhage AA. 2013. Regulation of fungal secondary metabolism. *Nat Rev Microbiol* 11:21–32. <http://dx.doi.org/10.1038/nrmicro2916>.
- Varga J, Houbbraken J, Van Der Lee HAL, Verweij PE, Samson RA. 2008. *Aspergillus calidoustus* sp. nov., causative agent of human infections previously assigned to *Aspergillus ustus*. *Eukaryot Cell* 7:630–638. <http://dx.doi.org/10.1128/EC.00425-07>.
- Egli A, Fuller J, Humar A, Lien D, Weinkauff J, Nador R, Kapasi A, Kumar D. 2012. Emergence of *Aspergillus calidoustus* infection in the era of posttransplantation azole prophylaxis. *Transplantation* 94:403–410. <http://dx.doi.org/10.1097/TP.0b013e31825992f0>.
- Brakhage AA, Van den Brulle J. 1995. Use of reporter genes to identify recessive *trans*-acting mutations specifically involved in the regulation of *Aspergillus nidulans* penicillin biosynthesis genes. *J Bacteriol* 177:2781–2788.
- 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 USA* 108:1513–1518. <http://dx.doi.org/10.1073/pnas.1017351108>.
- Haas BJ, Zeng W, Pearson MD, Cuomo CA, Wortman JR. 2011. Approaches to fungal genome annotation. *J Mycol* 2:118–141.
- 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>.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <http://dx.doi.org/10.1186/gb-2008-9-1-r7>.
- Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics* 4. <http://dx.doi.org/10.1002/0471250953.bi0406s35>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntentically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <http://dx.doi.org/10.1093/bioinformatics/btn013>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <http://dx.doi.org/10.1186/1471-2105-5-59>.
- Salamov AA, Solovvey VV. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res* 10:516–522. <http://dx.doi.org/10.1101/gr.10.4.516>.
- Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 31:5654–5666. <http://dx.doi.org/10.1093/nar/gkg770>.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <http://dx.doi.org/10.1038/nprot.2013.084>.
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol* 12:R22. <http://dx.doi.org/10.1186/gb-2011-12-3-r22>.
- Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, Miyasato SR, Binkley J, Orvis J, Shah P, Wymore F, Sherlock G, Wortman JR. 2014. The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-seq data to improve structural gene annotations. *Nucleic Acids Res* 42:D705–D710. <http://dx.doi.org/10.1093/nar/gkt1029>.
- Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <http://dx.doi.org/10.1186/1471-2105-6-31>.
- Keller O, Odonitz F, Stanke M, Kollmar M, Waack S. 2008. Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics* 9:278. <http://dx.doi.org/10.1186/1471-2105-9-278>.
- 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.
- 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>.
- 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>.
- 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>.
- 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>.