The draft genome sequence of the ascomycete fungus Penicillium subrubescens reveals a highly enriched content of plant biomass related CAZymes compared to related fungi

Peng, Mao

2017-03-20

Peng , M , Dilokpirnol , A , Mäkelä , M R , Hilden , K , Bervoets , S , Riley , R , Grigoriev , I V , Hainaut , M , Henrissat , B , de Vries , R P & Granchi , Z 2017 , ' The draft genome sequence of the ascomycete fungus Penicillium subrubescens reveals a highly enriched content of plant biomass related CAZymes compared to related fungi ' , Journal of Biotechnology , vol. 246 , pp. 1-3 . https://doi.org/10.1016/j.jbiotec.2017.02.012

http://hdl.handle.net/10138/231693
https://doi.org/10.1016/j.jbiotec.2017.02.012

Journal of Biotechnology 246 (2017) 1–3

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Short Genome Communications

The draft genome sequence of the ascomycete fungus *Penicillium subrubescens* reveals a highly enriched content of plant biomass related CAZymes compared to related fungi

Mao Peng[a], Adiphol Dilokpimol[a], Miia R. Mäkelä[a,b], Kristiina Hildén[b], Sander Bervoets[c], Robert Riley[d], Igor V. Grigoriev[d], Matthieu Hainaut[e,f], Bernard Henrissat[e,f,g], Ronald P. de Vries[a,b,*], Zoraide Granchi[c]

[a] *Fungal Physiology, Westerdijk Fungal Biodiversity Institute & Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands*
[b] *Division of Microbiology and Biotechnology, Department of Food and Environmental Sciences, University of Helsinki, Viikinkaari 9, Helsinki, Finland*
[c] *GenomeScan B.V., Plesmanlaan 1/D, 2333 BZ Leiden, The Netherlands*
[d] *US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, United States*
[e] *CNRS UMR 7257, Aix-Marseille University, 13288 Marseille, France*
[f] *INRA, USC 1408 AFMB, Marseille, France*
[g] *Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia*

ARTICLE INFO

Article history:
Received 26 December 2016
Received in revised form 10 February 2017
Accepted 13 February 2017
Available online 16 February 2017

ABSTRACT

Here we report the genome sequence of the ascomycete saprobic fungus *Penicillium subrubescens* FBCC1632/CBS132785 isolated from a Jerusalem artichoke field in Finland. The 39.75 Mb genome containing 14,188 gene models is highly similar for that reported for other *Penicillium* species, but contains a significantly higher number of putative carbohydrate active enzyme (CAZyme) encoding genes.

Penicillium subrubescens is a saprobic species and the strain sequenced here, FBCC1632/CBS132785, was originally isolated from soil of a Jerusalem artichoke field in Helsinki, Finland. It has been placed in section *Lanata-diversicata* of the genus *Penicillium*, which has several distinctive morphological features (Mansouri et al., 2013). The species has a high potential for the production of plant biomass degrading enzyme mixtures (Mäkelä et al., 2016).

The fungus was cultivated on complete medium (de Vries et al., 2004) and genomic DNA from 3-day old mycelium was extracted using CTAB-based extraction buffer (Hildén et al., 2005). The RNA was extracted using TRIzol reagent (Invitrogen/Thermo Fischer Scientific, Carlsbad, CA) and purified by NucleoSpin RNAII (Macherey-Nagel, Düren, Germany) from 3-day old cultures grown on wheat bran and sugar beet pulp in minimal medium (de Vries et al., 2004). Concentration and quality of the DNA were determined using the Life Technology Qubit and 0.6% agarose gel, respectively while the RNA samples quality was checked using Fragment Analyser (Advanced Analytical Technologies). Genome and transcriptome sequencing were performed in GenomeScan facilities.

The DNA was fragmented using the Covaris Focused-ultrasonicator. NEBNext® Ultra DNA Library Prep kit for Illumina (cat# NEB #E7370S/L) and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB #E7420S/L) were used according to manual for library preparation. Quality and yield after sample preparation was measured with Lab-on-a-Chip analysis or Fragment Analyzer. Clustering and DNA sequencing using the Illumina cBot and HiSeq 2500 was performed according to manufacturer's protocols using concentration of 8.0 pM of DNA and 16.0 pM of cDNA, standard Illumina primers and HiSeq control software HCS v2.2.58.

Image analysis, base calling, and quality check was performed with the Illumina data analysis pipeline RTA v1.18.64 and Bcl2fastq v1.8.4. Adapter trimming and quality filtering were performed using GenomeScan in-house tool FASTQFilter v2.05. Briefly: adapter sequences were removed from the read. Reads were then filtered and clipped: bases with phred scores below Q22 were removed and their reads were split; reads shorter than 36 bp were removed.

Abyss v1.3.7 (Simpson et al., 2009) was employed for the assembly, using k-mer length of 64. Scaffolds shorter than 500 bp were removed. The 39.75 Mb genome was obtained by assembly of 776 contigs (Table 1). The GC content was 49.21% as assessed by QUAST

* Corresponding author.
E-mail address: r.devries@westerdijkinstitute.nl (R.P. de Vries).

http://dx.doi.org/10.1016/j.jbiotec.2017.02.012
0168-1656/

Table 1
Genome features of *P. subrubescens* FBCC1632/CBS132785.

Features (# means the number)	
# of reads	32,833,084 (raw reads) 29,441,264 (filtered reads)
Paired-end read length (bp)	251
Genome assembly size (Mb)	39.75
# of contigs	1146
# of scaffolds	776
Scaffold N50 (bp)	322,783
# of gene models	14,188
# of exons per gene (average)	2.46
Mean protein length (aa)	416
GC content (%)	49.21
KEGG annotated (%)	12.20
KOG annotated (%)	45.94
InterPro annotated (%)	71.55
CAZymes	719

(Gurevich et al., 2013). The HMM-based algorithm Glimmer (version 3.02) (Majoros et al., 2004) was trained for gene finding using the genome of *Penicillium oxalicum* (Liu et al., 2013). Furthermore, an evidence-based method of gene finding was performed using the CodingQuarry (Testa et al., 2015) software tool and the mapped mRNA-Seq reads. The output of both methods was combined into a single gene model of 14,188 genes.

The CAZyme gene content of *P. subrubescens*, and of the compared fungal species, was determined as follows. All encoded protein models were compared using BLASTp (Altschul et al., 1997) to the proteins listed in the CAZy database (www.cazy.org, Lombard et al., 2014). All hits that gave an e-value >0.001 were kept for further analysis. Each model with >50% identity over the entire length of an entry in CAZy was directly assigned to the same family (or families in the case of modular proteins). Proteins with <50% identify to a protein in CAZy were manually aligned to known family members and searched for conserved features as the catalytic residues whenever known. Because a given percentage of sequence identity can result in widely different e-values (from non-significant to highly significant) depending on the length of the query sequence, no fixed e-value threshold was followed. In case of multimodular CAZymes, sequence alignments with isolated functional domains were performed (Cantarel et al., 2009). Related known activities information from derived from the CAZy database, crossed with literature data, were used to identify families likely involved in plant biomass degradation (Table 2, Suppl. Table 1).

This analysis revealed a significantly higher number of CAZy genes for *P. subrubescens* than that found in genomes of related species (Table 2, Suppl. Table 1). In particular, the number of glycoside hydrolases (GHs) is high and this remains the case when only CAZy genes related to plant biomass degradation are considered. The difference is especially significant when compared to the two other Penicillia, *P. rubens* and *P. chrysogenum*, as *P. subrubescens* contains approximately 50–70% more CAZy genes than those species. The number of CAZy genes for *P. subrubescens* is also much higher than recently reported for *Penicillium oxalicum* (Zhao et al., 2016) and *Talaromyces verruculoses* (Hu et al., 2016). Interestingly, the increase of GHs is not random, but affects specific families to different degrees. Particularly enriched plant biomass degradation related families are GH1, GH11, GH12, GH29, GH32, GH36, GH43, GH51, GH54, GH62 and GH67, and also carbohydrate esterase family CE8. Most of these families are related to hemicellulose or pectin degradation, with the exception of GH32, which contains inulin degrading enzymes. This fits well with the high inulinase activity that was previously reported for *P. subrubescens* (Mansouri et al., 2013). Recently, a comparison of *P. subrubescens* and *Aspergillus niger* during growth on wheat bran and sugar beet pulp revealed the ability of *P. subrubescens* to produce a set of plant biomass degrading enzymes with similar efficiency in plant biomass saccharification as *A. niger* (Mäkelä et al., 2016). The CAZy analysis of the *P. subrubescens* genome described here supports the high potential of this species as a producer of industrial enzyme mixtures.

This draft genome sequence of *P. subrubescens* FBCC1632/CBS132785 has been deposited at GenBank under accession number MNBE00000000. The version described in the paper is MNBE00000000. The BioProject in GenBank is PRJNA343459. The strain is available from the HAMBI-Fungal Biotechnology Culture Collection, University of Helsinki (e-mail: fbcc@helsinki.fi) and from the CBS collection (www.cbs.knaw.nl). The genome is also available through the JGI fungal genome portal MycoCosm (Grigoriev et al., 2014). This portal also provides more detailed information on the genome, such as orthologous groups, signal analysis and KEGG pathways.

Acknowledgment

This work was supported by the European Union, Grant agreement no: 613868 (OPTIBIOCAT).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2017.02.012.

Table 2
CAZyme content of selected fungal genomes in number of genes per group. PBD = plant biomass degradation related genes, GH = glycoside hydrolase, PL = polysaccharide lyase, CE = carbohydrate esterase, AA = auxiliary activity, GT = glycosyl transferase, CBM = carbohydrate binding module, EXP = expansin. CAZy numbers for all species were analysed using the CAZy annotation pipeline and currently available public version of the genomes.

Species	Strain	Total CAZy	GH		PL		CE		AA		GT	CBM	EXP
			Total	PBD	Total	PBD	Total	PBD	Total	PBD			
Penicillium subrubescens	FCBB1632/CBS132785	719	410	241	9	9	38	38	63	60	107	85	7
Penicillium rubens	Wisconsin 54-1255	426	222	120	9	9	20	20	22	13	101	51	1
Penicillium chrysogenum	unknown	481	234	125	9	9	20	20	50	45	110	56	2
Talaromyces stipitatus	ATCC 10500	514	271	136	2	0	17	18	47	41	105	65	7
Aspergillus niger	NRRL3	542	252	137	9	9	22	22	65	59	119	72	3
Aspergillus oryzae	RIB40	600	304	174	23	20	27	26	69	62	119	54	4
Aspergillus nidulans	FGSC A4	572	275	161	23	21	28	28	57	54	97	90	2
Trichoderma reesei	QM6a	410	200	77	5	0	16	15	32	27	92	58	7
Myceliophthora thermophila	ATCC 42464	419	197	100	8	7	25	25	52	43	84	51	2
Neurospora crassa	OR74A	416	182	77	4	3	22	22	51	42	86	68	3
Podospora anserina	S mat+	590	212	104	7	7	40	40	109	96	93	123	6

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, D233–D238.

de Vries, R.P., Burgers, K., van de Vondervoort, P.J.I., Frisvad, J.C., Samson, R.A., Visser, J., 2004. A new black Aspergillus species, A. vadensis, is a promising host for homologous and heterologous protein production. Appl. Environ. Microbiol. 70, 3954–3959.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699–704.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.

Hildén, K., Martínez, A.T., Hatakka, A., Lundell, T., 2005. The two manganese peroxidases Pr-MnP2 and Pr-MnP3 of *Phlebia radiata* a lignin-degrading basidiomycete, are phylogenetically and structurally divergent. Fungal Genet. Biol. 42, 403–419.

Hu, L., Taujale, R., Liu, F., Song, J., Yin, Q., Zhang, Y., Guo, J., Yin, Y., 2016. Draft genome sequence of *Talaromyces verruculosus* (*Penicillium verruculosum*) strain TS63-9, a fungus with great potential for industrial production of polysaccharide-degrading enzymes. J. Biotechnol. 219, 5–6.

Liu, G., Zhang, L., Wei, X., Zou, G., Qin, Y., Ma, L., Li, J., Zheng, H., Wang, S., Wang, C., Xun, L., Zhao, G.P., Zhou, Z., Qu, Y., 2013. Genomic and secretomic analyses reveal unique features of the lignocellulolytic enzyme system of *Penicillium decumbens*. PLoS One 8, e55185.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495.

Mäkelä, M.R., Mansouri, S., Wiebenga, A., Rytioja, J., de Vries, R.P., Hildén, K., 2016. *Penicillium subrubescens* is a promising alternative for *Aspergillus niger* in enzymatic plant biomass saccharification. N. Biotechnol. 33, 834–841.

Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879.

Mansouri, S., Houbraken, J., Samson, R.A., Frisvad, J.C., Christensen, M., Tuthill, D.E., Koutaniemi, S., Hatakka, A., Lankinen, P., 2013. *Penicillium subrubescens*, a new species efficiently producing inulinase. Antonie Van Leeuwenhoek 103, 1343–1357.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117–1123.

Testa, A.C., Hane, J.K., Ellwood, S.R., Oliver, R.P., 2015. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genom. 16, 170.

Zhao, S., Yan, Y.S., He, Q.P., Yang, L., Yin, X., Li, C.X., Mao, L.C., Liao, L.S., Huang, J.Q., Xie, S.B., Nong, Q.D., Zhang, Z., Jing, L., Xiong, Y.R., Duan, C.J., Liu, J.L., Feng, J.X., 2016. Comparative genomic, transcriptomic and secretomic profiling of *Penicillium oxalicum* HP 7-1 and its cellulase and xylanase hyper-producing mutant EU2106, and identification of two novel regulatory genes of cellulase and xylanase gene expression. Biotechnol. Biofuels 9, 203.