Genome sequence of Rhizomucor pusillus FCH 5.7, a thermophilic zygomycete involved in plant biomass degradation harbouring putative GH9 endoglucanases

Downloaded from: https://research.chalmers.se, 2020-03-06 04:51 UTC

Citation for the original published paper (version of record)
Genome sequence of Rhizomucor pusillus FCH 5.7, a thermophilic zygomycete involved in plant biomass degradation harbouring putative GH9 endoglucanases
Biotechnology Reports, 20
http://dx.doi.org/10.1016/j.btre.2018.e00279

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)
Short Communication

Genome sequence of *Rhizomucor pusillus* FCH 5.7, a thermophilic zygomycete involved in plant biomass degradation harbouring putative GH9 endoglucanases

Silvia Hüttner\(^{a,b,*}\), Zoraide Granchi\(^{c}\), Thanh Thuy Nguyen\(^{d}\), Sake van Pelt\(^{c}\), Johan Larssrink\(^{a,b}\), Vu Nguyen Thanh\(^{d}\), Lisbeth Olsson\(^{a,b}\)

\(^{a}\)Department of Biology and Biological Engineering, Division of Industrial Biotechnology, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
\(^{b}\)Wallenberg Wood Science Center, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
\(^{c}\)GenomeScan B.V., Plasmanlaan 1D/0, 2333 BZ Leiden, The Netherlands
\(^{d}\)Centre for Industrial Microbiology, Food Industries Research Institute, Thanh Xuan, Ha Noi, Viet Nam

1. Introduction

Fast growth and rapid spore production mark the thermophilic zygomycete *Rhizomucor pusillus*, a member of the Mucorales, as a mycophora pioneer species in compost [1,2]. A number of its encoded enzymes have been produced and characterised, such as polygalacturonase, glycoamylase, phytase and glucanase [3–6], and particularly the high temperature optima of *R. pusillus* enzymes of around 60–70 °C [5,6] make the fungus an interesting source of biocatalysts. A lipase of a close relative, *Rhizomucor miehei*, is commercially available and used industrially (Palatase\(^{a}\), Novozymes) [7]. Therefore, *R. pusillus* represents a promising resource for the exploration of useful enzymes. The *R. pusillus* strain FCH 5.7 described in this study was isolated from 3 month old compost in Ly Nhan, Ha Nam province, Vietnam (geographical location 20°33’06.2” N, 106°04’11.8” E). Genome sequencing was performed to enable a systematic exploration, particularly of the strain’s carbohydrate-active enzymes (CAZymes) [8], as well as to set the basis for investigating its growth strategies at higher temperatures.

2. Materials & methods

After isolation, *R. pusillus* FCH 5.7 was cultivated on potato dextrose agar (PDA) at 50 °C and the species identified by internal transcribed spacer (ITS) sequencing, as described previously [9]. For whole genome sequencing, 5-day-old mycelium from PDA plates was used for the inoculation of basal liquid medium (4 g L\(^{-1}\) KH\(_2\)PO\(_4\), 13.6 g L\(^{-1}\) (NH\(_4\))\(_2\)SO\(_4\), 0.8 g L\(^{-1}\)CaCl\(_2\)-2H\(_2\)O, 0.6 g L\(^{-1}\) MgSO\(_4\)-7H\(_2\)O, 6 g L\(^{-1}\) Bacto peptone, 10 mg L\(^{-1}\) FeSO\(_4\)-7H\(_2\)O, 3.2 mg L\(^{-1}\) MnSO\(_4\)-H\(_2\)O, 2.8 mg L\(^{-1}\) ZnSO\(_4\)-7H\(_2\)O, 4 mg L\(^{-1}\) CoCl\(_2\)-6H\(_2\)O, 200 mL L\(^{-1}\) Tween 80, pH 5.8) containing 20 g L\(^{-1}\) glucose as carbon source. Cultivation was carried out in 125 mL liquid medium in baffled Erlenmeyer flasks (500 mL; 50 °C, 48 h, shaking at 250 rpm). The mycelium was harvested after two days of cultivation and DNA was extracted with cetyltrimethyl ammonium bromide (CTAB) buffer (2% CTAB, 100 mM TrisHCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl) and purified from the supernatant by a combination of phenol-chloroform extraction and isopropanol precipitation and DNeasy Plant Mini Kit (Qiagen), as described...
previously [9]. Quality of the purified DNA was verified by agarose gel electrophoresis, Nanodrop (Thermo Scientific) and Qubit (Life Technologies) before genome sequencing. RNA was extracted from cultures growing in basal liquid medium, supplemented with 1% glucose, wheat bran or beechwood xylan, respectively, after 4 h and 48 h, using Trizol (Invitrogen) and chloroform, according to the manufacturer’s instructions. Enzymatic activities (xylanase, endoglucanase) were determined in supernatant samples from the same cultivations through measurement of produced reducing ends, using the 3,5-dinitrosalicylic acid DNS method [10].

For genome sequencing, the NEBNext Ultra DNA Library Prep kit for Illumina (New England Biolabs) was used to process the DNA samples according to the manufacturer’s protocol. After fragmentation of DNA with a Covaris ultrasonicator (Thermo Scientific), quality and yield were measured with a Bioanalyzer (Agilent Technologies). Clustering and DNA sequencing with the Illumina cBot and HiSeq 2500 were performed with 8.0 pM of DNA, standard Illumina primers and HiSeq control software HCS v2.2.58. Image analysis, base calling, and quality check were done with the Illumina data analysis pipeline RTA v1.18.64 and Bcl2fastq v1.8.4. The GenomeScan in-house tool FASTQFilter v2.05 was used for adapter trimming and quality filtering of the 250 bp paired-end reads. The short-read genome assembler Abyss v1.3.7 [11] with a k-mer length of 64 was used for assembly. Scaffolds shorter than 500 bp were removed.

For RNA sequencing, total RNA was further prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) with oligo-dT magnetic beads for mRNA enrichment. Sequencing of the resulting cDNA (16 pM) was done with an Illumina cBot and HiSeq 2500 instrument. The newly assembled genome was used as a reference to map the mRNA-Seq reads using the packages Tophat (v2.0.14, Linux_x86_64) and Bowtie (v2-2.1.0) with a default mismatch rate of 2%. Gene finding was performed combining two approaches. The HMM-based algorithm Glimmer v.3.02 [12] was trained using the genome of *Mucor circinelloides* (downloaded from the Joint Genome Institute (JGI), http://genome.jgi.doe.gov/). In addition, an evidence-based method of gene finding was performed using mapped mRNA-Seq reads and the software tool CodingQuarry [13]. Combining the two methods, gene models for 10,898 genes were obtained. For annotation, a BLASTp search (version 2.2.28+) (http://blast.ncbi.nlm.nih.gov) was performed on the UniProtKB/SwissProt database with default parameters. For comparison to other species, publicly accessible genomes were downloaded from JGI Mycococosm (https://genome.jgi.doe.gov/programs/fungi/index.jsf, Supplemental File 1).

The CAZyme contents of *R. pusillus* FCH 5.7 and the fungal species included for comparison were determined by identifying genes containing CAZyme domains using the dbCAN2 meta server (cys.bios.niu.edu/dbCAN2). Only CAZyme domains predicted by at least two of the three algorithms (DIAMOND, HMMER and Hotpep) employed by dbCAN2 were kept. It should be noted that the stringency of this approach led to slightly different numbers of identified CAZymes than has previously been reported for the respective species.

### 3. Results & discussion

The 25.59 Mb genome of *R. pusillus* FCH 5.7 was obtained with 294-fold coverage (Table 1). The whole genome assembly has been submitted to the European Nucleotide Archive and deposited at DDBJ/EMBL/GenBank under the assembly accession number PWWN000000000.

The analysis of CAZyme-encoding genes showed that *R. pusillus* FCH 5.7 possesses a similar number of total CAZymes as other zygomycetes, such as *Hesseltinella vesiculosa* and *Phycomyces blakesleeanus*, but considerably fewer than *Lichtheimia corymbifera* and *Rhizopus oryzae* (Table 2), which could be due to the reported gene expansion and genome duplication in *L. corymbifera* and *R. oryzae*, respectively [14,15]. The difference was particularly obvious for carbohydrate esterases (CE) and glycoside transfersases (GT), two CAZy enzyme classes that are known to be highly abundant in *R. oryzae* [16]. Compared to basidiomycetes, zygomycetes generally have higher numbers of GTs, and this was also true for *R. pusillus*, which harbours 77 CAZy domains across 27 GT families in its genome. The number of auxiliary activity (AA) enzymes was relatively small in the investigated zygomycetes, and *R. pusillus* FCH 5.7 was found to only possess members of AA families 1, 5 and 6, as well as one member in family 12. None of the zygomycetes used for comparison contained AA9 lytic polysaccharide monoxygenases (LPMOs), which are abundant in ascomycetes and basidiomycetes and involved in cellulose degradation (Supplemental File 1; [17]).

Growth on plates containing basal medium and 1% beechwood xylan was very poor, and *R. pusillus* FCH 5.7 grew much better on mono- and disaccharides (glucose, galactose, mannose, cellulose), as well as starch (Fig. 1A). *R. pusillus* FCH 5.7 was not able to grow efficiently on cellulose (Avicel), which indicates a lack of endoglucanases and/or cellobiolydrolases to break down this polymer. Indeed, no putative GH6 or GH7 cellobiolydrolases could be found in the *R. pusillus* FCH 5.7 genome. The ability to degrade chitin was indicated by the presence of many chitinases (GH18) and chitin deacytelylases (CE4) in the genome, and was corroborated by the ability of *R. pusillus* FCH 5.7 to grow on chitosan (Fig. 1A, Supplemental File 1). In contrast to a previous study [4], *R. pusillus* FCH 5.7 grew quite poorly on pectin, and many genes typically involved in pectin degradation, such as pectin and pectate lyases (PL1, 3, 4), were found missing in its genome. We did, however, detect sequences of three GH28s that encode putative polygalacturonases. Similar to *R. miehei* [18], a large number of lipases was found in the genome of *R. pusillus* FCH 5.7, with in total 80 putative lipases and phospholipases. Despite the poor growth on plates containing Avicel or xylan, xylanase and low levels of endoglucanase (CMCase) activities were

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Genome features of <em>R. pusillus</em> FCH 5.7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome assembly size (Mb)</td>
<td>25.59</td>
</tr>
<tr>
<td>Genome coverage (x times)</td>
<td>294</td>
</tr>
<tr>
<td>Number of scaffolds (≥500 bp)</td>
<td>618</td>
</tr>
<tr>
<td>Contig N50 (bp)</td>
<td>102,680</td>
</tr>
<tr>
<td>Number of scaffolds &gt; N50</td>
<td>77</td>
</tr>
<tr>
<td>Max scaffold size (bp)</td>
<td>534,899</td>
</tr>
<tr>
<td>Number of protein coding genes</td>
<td>10,898</td>
</tr>
<tr>
<td>Mean gene length (bp)</td>
<td>1,359</td>
</tr>
<tr>
<td>GC content, overall (%)</td>
<td>44.95</td>
</tr>
<tr>
<td>GC content, coding (%)</td>
<td>48.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Comparison of the numbers of CAZymes in <em>R. pusillus</em> FCH 5.7 with those in other fungi.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZYG</td>
<td>Rhizomucor pusillus</td>
</tr>
<tr>
<td>Abisidium repens</td>
<td>236</td>
</tr>
<tr>
<td>Hesseltinella vesiculosa</td>
<td>183</td>
</tr>
<tr>
<td>Lichtheimia corymbifera</td>
<td>261</td>
</tr>
<tr>
<td>Mucor circinelloides</td>
<td>211</td>
</tr>
<tr>
<td>Phycycomyces blakesleeanus</td>
<td>195</td>
</tr>
<tr>
<td>Rhizopus microsporus</td>
<td>201</td>
</tr>
<tr>
<td>Rhizopus oryzae</td>
<td>263</td>
</tr>
<tr>
<td>Syncephalastrum racemosum</td>
<td>239</td>
</tr>
<tr>
<td>ASC</td>
<td>Theliacea terrestres</td>
</tr>
<tr>
<td>BAS</td>
<td>Schizopyllum commune</td>
</tr>
</tbody>
</table>
detected in liquid cultures of \textit{R. pusillus} FCH 5.7 grown on wheat bran or beechwood xylan as carbon sources (Fig. 1A). Cultivation on wheat bran and beechwood xylan resulted in almost identical levels of CMCase activity. Xylanase activity, on the other hand, was much higher during growth on xylan compared to growth on wheat bran. This indicates the presence of xylanases that are induced more by hardwood glucuronoxylans, as found in beechwood, rather than the grass-specific arabinoxylans contained in wheat bran.

Several putative endoglucanases from family GH45 were detected in the genome, which could explain the low enzyme activity observed on CMC (Fig. 1B). Interestingly, \textit{R. pusillus} FCH 5.7 also encodes two GH9 members, which is a family containing endoglucanases typically found in bacteria, plants and occasionally animals [19,20], but completely absent in ascomycetes fungi. We compared the genomes of a selection of fungi to determine how widespread the occurrence of GH9 genes is. Many of the basidiomycetes in our selection were shown to encode a single putative GH9 protein, while up to four were detected in zygomycetes (Fig. 2). Zygomycete GH9s are, to our knowledge, completely unexplored proteins and the function of fungal GH9s and the importance for, for instance, growth on polysaccharides is unclear and has not been described yet. One study indicated an involvement in degradation of crystalline cellulose, as a putative GH9 gene was found to be upregulated in the basidiomycete \textit{P. chrysosporium} grown on cellulose [21], but conclusive functional evidence is still missing.

The lifestyle and habitat, and therefore the repertoire of CAZymes of zygomycetes is quite different to ascomycetes and basidiomycetes. The enzyme system of \textit{R. pusillus} FCH 5.7 seems to be better suited to consume easily accessible, simple sugars found in plant biomass [22] as well as lipid-containing compounds. As seen in the growth studies (Fig. 1A) and also shown for \textit{R. oryzae}, zygomycetes are in general unable to degrade complex plant cell wall polysaccharides [16]. The lack of genes encoding enzymes of the CAZy families GH6, GH7 and AA9 in the \textit{R. pusillus} genome further supports this statement. However, the low endoglucanase and xylanase activities detected in \textit{R. pusillus} FCH 5.7 culture supernatants, as well as the presence of putative xylanases and endoglucanases (GH9, GH45) in the genome indicate that this fungus may be able to degrade oligo- and/or polysaccharide to a certain extent.

3.1. Nucleotide sequence accession number(s)

The \textit{R. pusillus} FCH 5.7 whole genome sequence has been submitted to GenBank and deposited at DDBJ/EMBL/GenBank under the accession No. FWN00000000.
Conflict of interest statement

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was financially supported by the European Union OPTIBIOCAT project (FP7 KBBE. 2013.3.3-04; grant agreement no. 613868), by the Swedish Research Council (Vetenskapsrådet) grant no. 2014-3523, and by the Wallenberg Wood Science Center, granted by the Wallenberg Foundation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.btre.2018.e00279.

References