Gene 519 (2013) 360–366

Contents lists available at SciVerse ScienceDirect

Gene

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

A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost

Yeh Yi-Fang [a,1], Chang Samuel Chia-yu [a,1], Kuo Hsion-Wen [a], Tong Chi-Gong [a], Yu Su-May [b,*], Ho Tuan-Hua David [c,d,**]

[a] Biotechnology Center in Southern Taiwan, Academia Sinica, No. 59, Xiraya Blvd., Xinshi Dist., Tainan 74145, Taiwan, ROC
[b] Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 115, Taiwan, ROC
[c] Institute of Plant and Microbial Biology, Academia Sinica, Nankang, Taipei 115, Taiwan, ROC
[d] Department of Biology, Washington University, St. Louis, MO 63130, USA

ARTICLE INFO

Article history:
Accepted 31 July 2012
Available online 9 August 2012

Keywords:
Metagenomic library
Rice straw composts
Functional screening
Endoglucanase
Industrial application

ABSTRACT

Over the years, culturable cellulase-producing microorganisms have been isolated from a variety of sources and genes of cellulolytic enzymes have been cloned. Then again, the "great plate count anomaly" phenomenon necessitates a culture-independent metagenomic approach for the isolation of cellulolytic genes from microorganisms in their natural environment. We have constructed a metagenomic library derived from rice straw composts. Of 2739 clones screened, a lambda clone carrying a 12 kb genomic fragment was able to yield a clear zone on an agar plate containing carboxymethyl cellulose (CMC). A 4.7 kb subclone, generated by restriction enzyme digestion, was found to harbor a GH12 cellulase gene, *RSC-EG1*, encoding 464 amino acids along with two other ORFs. The identified cellulolytic gene showed more than 70% similarity on the amino acid level with cellulase from *Micromonospora aurantiaca* and *Thermobispora sp.* Interestingly, RSC-EG1 contains a stretch of approximately 86 amino acids not present in either of these close relatives. Our results demonstrated that RSC-EG1, stable over a wide temperature and pH range, is a novel endoglucanase, and provided the first example of metagenomics approach to isolate cellulolytic gene from rice straw composts.

1. Introduction

The imminent arrival of peak oil (Kerr, 2011) underscores the importance of developing renewable fuel alternatives. The application of metabolic engineering to establish microbes that produce advanced biofuels from diverse lignocellulosic materials provides one such sustainable alternative (Peralta-Yahya and Keasling, 2010). A distinctive feature of cellulose in plant biomass is its crystalline structure in which individual microfibrils are packed so tightly that molecules such as enzymes cannot penetrate (Lynd et al., 2002). The enzymatic conversion of cellulose into glucose is therefore challenging due to the physical nature of the substrate.

As plant biomass is naturally recycled in the environment, a diverse spectrum of lignocellulolytic microorganisms must exist that hydrolyze and metabolize cellulose efficiently (Bouws et al., 2008; McCarthy and Williams, 1992). Over the years, bioenergy researchers have isolated cellulase-producing microorganisms from a wide variety of sources and cloned genes of cellulolytic enzymes (Adney et al., 1991; Kumar et al., 2008; Teunissen and Op den Camp, 1993). However, the microbial population retrievable by standard cultivation techniques covers only a small fraction of the diversity present in environmental samples (Amann et al., 1995; Torsvik et al., 1998). To circumvent the discrepancy between culturable and *in situ* microbial diversity (Staley and Konopka, 1985), Pace et al. (1986) proposed the direct extraction of microbial DNA from the environment in order to obtain an unbiased genomic representation. The analysis of natural microbial communities by the direct extraction and cloning of DNA was later termed metagenomics (Handelsman et al., 1998).

Since then, a remarkable number of publications have reported the isolation of metagenome-derived cellulases (Grant et al., 2004; Rees et al., 2003; Voget et al., 2006; Warnecke et al., 2007). While functional screenings of metagenomic libraries from extreme environments have received a great deal of attention from researchers (Grant et al., 2004; Rees et al., 2003), non-extreme environments with significant genetic diversity also possess a wide range of cellulases that are suitable for industrial applications (Voget et al., 2006).

Abbreviations: BLAST, basic local alignment search tool; bp, base pair; CBM, cellulose-binding moule; DNA, deoxyribonucleic acid; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetate; GH, glycosyl hydrolase; kb, kilobase; kDa, kilodalton; PCR, polymerase chain reaction; rpm, revolutions per minute; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; std, standard; t-RNA, transfer RNA; v/v, volume/volume; wt, weight; w/v, weight-in-volume.

* Corresponding author. Tel.: +886 2 2788 2695; fax: +886 2 2782 6085.
** Corresponding author. Tel.: +886 2 2787 1079; fax: +886 2 2782 1605.
E-mail addresses: sumay@imb.sinica.edu.tw (S.-M. Yu), ho@biology2.wustl.edu, tho@gate.sinica.edu.tw (T.-H.D. Ho).
[1] Co-first author.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.07.076

Given that rice straw is one of the most abundant bioenergy feedstocks in the world (Kim and Dale, 2004), the discovery of cellulases for the conversion of this agricultural waste into glucose can have considerable social and economic benefits. In this work, we demonstrate the utility of metagenomics coupled with functional selections for the discovery of a novel cellulolytic gene from the rice straw compost microbiome. The major steps involved extraction of metagenomic DNA from rice straw compost, construction of a genomic library, and selection of clones able to yield a clear zone on CMC-agar plates. We identified an endoglucanase, Rice Straw Compost-Endoglucanase 1 (RSC-EG1), sharing similarity with cellulases of *Micromonospora aurantiaca*, which is stable over a wide temperature and pH range.

2. Materials and methods

2.1. Extraction and purification of DNA from environmental samples

Rice straw composts [chicken manure, rice seed hulls and rice straw at a ratio of 3:12:6 (wt.%)], where environmental samples were collected, were established at temperatures ranging from 55 to 70 °C at the Asian Vegetable Research and Development Center, Tainan, Taiwan. For DNA extraction, the direct lysis method was used (Zhou et al., 1996) with minor modifications. Briefly, 1 g of sample was homogenized by vortexting in 2.6 ml extraction buffer [100 mM Tris–HCl, pH 8; 100 mM sodium EDTA, pH 8; 100 mM sodium phosphate, pH 8; 1.5 M NaCl, 1% hexadecyltrimethylammonium bromide (CTAB)]. Three freeze–thaw cycles were carried out in liquid nitrogen and at 65 °C. After adding 50 μl of proteinase K (20 mg/ml), samples were incubated at 37 °C for 30 min with continuous shaking at 120 rpm. Subsequently, 300μl of 20% (w/v) SDS was added and incubated at 65 °C for 2 h with gentle shaking every 15–20 min. The supernatants were collected after centrifugation at 4000 *g* for 10 min and the resulting pellets were re-extracted in 2 ml extraction buffer at 65 °C for 10 min. The combined supernatants were then mixed with 1/10 volume of 10% (w/v) CTAB, which is able to complex with polysaccharides, denatured proteins and cell debris, keeping nucleic acid in solution for further use (Murray and Thompson, 1980), and centrifuged again. The resulting supernatant solutions were extracted with chloroform–isoamylalcohol (24:1, v/v), and DNA was precipitated with isopropanol, washed with 70% (v/v) ethanol, dried and resuspended in 100 μl of 10 mM Tris–HCl,

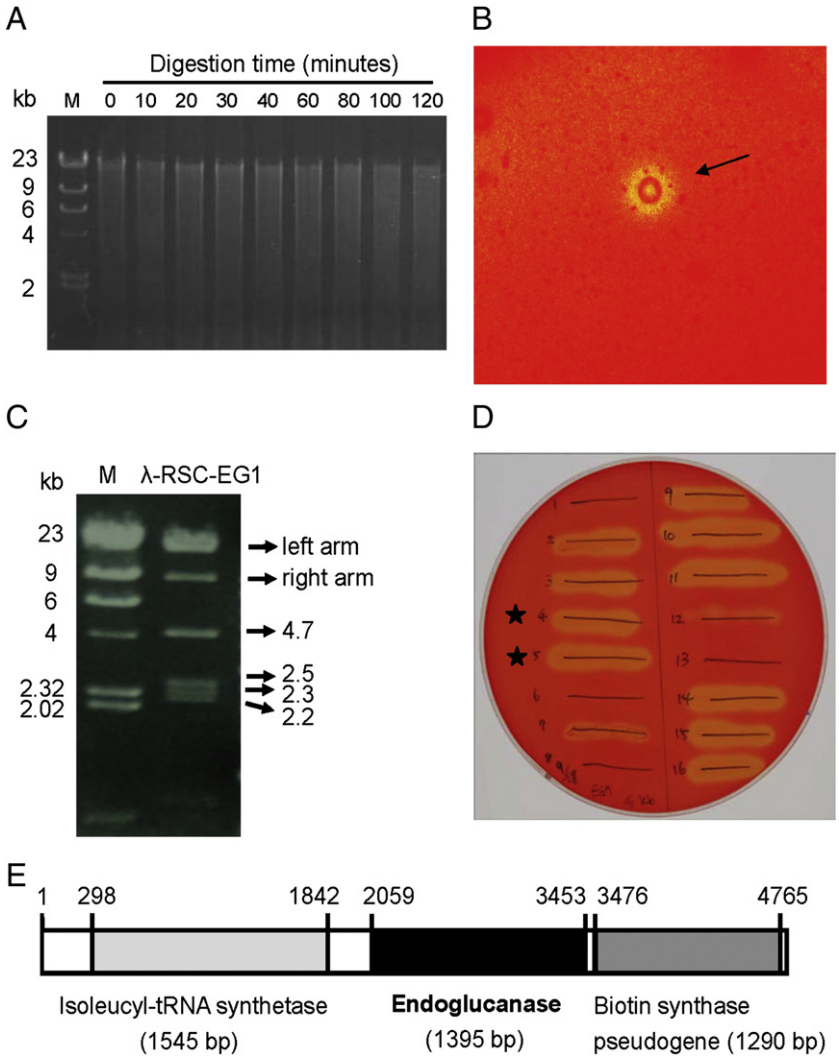

Fig. 1. Cloning of a novel endoglucanase gene from rice straw compost using a functional metagenomic approach. CMC plate assay is based on the interaction between Congo Red and CMC (Teather and Wood, 1982). (A) DNA from environmental sample was purified and partially digested with *Sau*3AI. (B) A lambda clone, out of 2739 clones, carrying a 12 kb DNA fragment was able to yield a clear zone on CMC-agar plates. Since Congo Red can be removed from CMC degraded by endoglucanases, clearing zones (black arrowhead) correspond to areas of endoglucanase activity. (C) The *Bam*H1 restriction enzyme digestion of λ-RSC-EG1 generated six fragments. Top two bands are from the lambda cloning vector and bottom four bands are sequences of interest. (D) The 4.7 kb fragment was the one conferring production of endoglucanase. Sixteen subclones were plated in liner fashion to demonstrate endoglucanase activity *in situ*, only 4 clones failed to show recombinant enzyme activity. (E) Two clones (★) from the CMC-agar plate in (D) were chosen for DNA sequencing. 4.7 kb fragment contains three ORFs: isoleucyl-tRNA-synthetase, GH12 endoglucanase, and biotin synthase pseudogene. The novel endoglucanase gene is 1395 bp long (464 amino acids) and was named *RSC-EG1*.