ORIGINAL PAPER

Identification and characterization of CbeI, a novel thermostable restriction enzyme from *Caldicellulosiruptor bescii* DSM 6725 and a member of a new subfamily of HaeIII-like enzymes

Dae-Hwan Chung · Jennifer R. Huddleston ·
Joel Farkas · Janet Westpheling

Abstract   Potent HaeIII-like DNA restriction activity was detected in cell-free extracts of *Caldicellulosiruptor bescii* DSM 6725 using plasmid DNA isolated from *Escherichia coli* as substrate. Incubation of the plasmid DNA in vitro with HaeIII methyltransferase protected it from cleavage by HaeIII nuclease as well as cell-free extracts of *C. bescii*. The gene encoding the putative restriction enzyme was cloned and expressed in *E. coli* with a His-tag at the C-terminus. The purified protein was 38 kDa as predicted by the 981-bp nucleic acid sequence, was optimally active at temperatures between 75°C and 85°C, and was stable for more than 1 week when stored at 35°C. The cleavage sequence was determined to be 5′-GG/CC-3′, indicating that CbeI is an isoschizomer of HaeIII. A search of the *C. bescii* genome sequence revealed the presence of both a HaeIII-like restriction endonuclease (Athe 2438) and DNA methyltransferase (Athe 2437). Preliminary analysis of other *Caldicellulosiruptor* species suggested that this restriction/modification activity is widespread in this genus. A phylogenetic analysis based on sequence alignment and conserved motif searches identified features of CbeI distinct from other members of this group and classified CbeI as a member of a novel subfamily of HaeIII-like enzymes.

Introduction

Caldicellulosiruptor bescii DSM 6725 (formerly *Anaerocellum thermophilum* [31]) grows at up to 90°C and is the most thermophilic cellulolytic bacterium known. This obligate anaerobe is capable of degrading lignocellulosic biomass including hardwood (poplar) and grasses with both low lignin (Napier grass, Bermuda grass) and high lignin (switchgrass) content without chemical pretreatment [30]. When grown on crystalline cellulose, it produces lactate, ethanol, acetate, H_2, and CO_2 [28]. Its genome includes sequences encoding cellulases, glycoside hydrolases, pullulanases, and transporters that are important in biomass deconstruction [11]. This variety of cellulolytic enzymes and end products, in combination with an optimal growth temperature near 80°C, make *C. bescii* an important microorganism not only in the study of biomass deconstruction, but also for industrial development of ethanol and other biofuels. This genus has many advantages for consolidated bioprocessing (CBP) and offers the possibility for production of bioenergy and bioproducts from lignocellulosic biomass by a *single* organism in a *single-step* fermentation [14].

The ability to genetically manipulate *Caldicellulosiruptor* species is a prerequisite for their use in CBP, and in our efforts to develop a method of DNA transformation we discovered several potent restriction activities, one of which we describe in this report. The activity of host restriction enzymes is a major barrier to introduction of DNA into cells, and identifying restriction systems and

D.-H. Chung · J. R. Huddleston ·
J. Farkas · J. Westpheling (✉)
Department of Genetics, University of Georgia, Athens,
GA 30602, USA
e-mail: janwest@uga.edu

D.-H. Chung · J. R. Huddleston · J. Farkas · J. Westpheling
BioEnergy Science Center, Biosciences Division, Oak Ridge
National Laboratory, Oak Ridge, TN 37831, USA

overcoming them through a variety of methods has allowed genetic systems to be developed in previously nontransformable bacteria. Often this is accomplished by in vitro methylation or by in vivo methylation systems in *E. coli*; for example, in the thermophile *Bacillus methanolicus*, plasmids are engineered with fewer BmeI recognition sites and prepared in a *dam*+ *E. coli* strain in order for transformation to occur [6]. In *Prevotella* species and *Helicobacter pylori*, plasmid DNA is methylated by cell-free extracts (CFEs) to achieve transformation by electroporation [1, 7]. In *Clostridium difficile*, plasmids are constructed that lack CdiI and Sau96I recognition sequences [23]. *Clostridium perfringens* type B transformation can only occur when the transforming plasmid DNA is isolated from a *dam*+ *dcm*+ strain of *E. coli* [5]. *Clostridium cellulolyticum* can only be transformed if the plasmid DNA is protected from CbeI cleavage using either in vitro or in vivo methylation [10].

In our efforts to develop efficient methods for DNA transformation of *C. bescii*, we identified a potent thermostable type II restriction endonuclease that is an isoschizomer of HaeIII [16]. HaeIII-like enzymes are a diverse group of proteins with distinct catalytic domains that have in common the ability to cleave the same DNA sequence. The prototype of this group, HaeIII, was first identified in *Haemophilus aegyptius* in 1972 [16]. This enzyme recognizes 5′-GGCC-3′ and cleaves the DNA between the G and C, leaving a blunt end [4]. HaeIII-like restriction activity is widespread in bacteria and archaea [25], allowing efficient restriction of foreign DNA. Four-base cutters such as HaeIII are a particularly serious problem in DNA transfer, since the occurrence of their cleavage sites is more frequent than longer recognition sequences. The enzyme identified from *C. bescii* was named CbeI, and its cleavage activity, temperature profile, and thermostability are described. This is the first investigation of a restriction-modification system in any species of *Caldicellulosiruptor*, and bioinformatic analysis of CbeI and other HaeIII isoschizomers revealed a previously unidentified subfamily of this group of restriction endonucleases. We suggest that the work described herein will facilitate the study of the nature of restriction-modification systems and assist in the establishment of genetic methods for this important group of organisms.

Materials and methods

Strains and growth conditions

All *Caldicellulosiruptor* species were grown in DSMZ 516 medium [28] with the following modifications. The mineral solution contained the following (per liter): NH_4Cl, 0.25 g;

KH_2PO_4, 0.33 g; KCl, 0.33 g; $MgCl_2·6H_2O$, 0.33 g; $CaCl_2·2H_2O$, 0.33 g; yeast extract, 0.5 g; casein hydrolysate (enzymatic; U.S. Biochemicals), 5 g; cellobiose, 5 g; resazurin, 0.25 mg; vitamin solution, 2 ml; trace minerals solution, 1 ml; amino acid solution, 40 ml. The vitamin solution contained the following (per liter): biotin, 10 mg; folic acid, 10 mg; pyridoxine–HCl, 50 mg; thiamine-HCl, 25 mg; riboflavin, 25 mg; nicotinic acid, 25 mg; calcium pantothenate, 25 mg; vitamin B_{12}, 0.5 mg; *p*-aminobenzoic acid, 25 mg; lipoic acid, 25 mg. The trace element solution contained the following (per liter): HCl (25%: 7.7 M), 1.0 ml; $FeCl_3·4H_2O$, 2 g; $ZnCl_2$, 50 mg; $MnCl_2·4H_2O$, 50 mg; H_3BO_3, 50 mg; $CoCl_2·6H_2O$, 50 mg; $CuCl_2·2H_2O$, 30 mg; $NiCl_2·6H_2O$, 50 mg; Na_4EDTA (tetrasodium salt), 50 mg; $(NH_4)_2MoO_4$, 50 mg; $AlK(SO_4)_2·12H_2O$, 50 mg. The amino acid solution contained the following (per liter): L-alanine, 1.9 g; L-arginine, 3.1 g; L-asparagine, 2.5 g; L-aspartic acid, 1.2; L-glutamic acid, 5.0 g; L-glutamine, 1.2 g; glycine, 5.0 g; L-histidine, 2.5 g; L-isoleucine, 2.5 g; L-leucine, 2.5 g; L-lysine, 2.5 g; L-methionine, 1.9 g; L-phenylalanine, 1.9 g; L-proline, 3.1 g; L-serine, 1.9 g; L-threonine, 2.5 g; L-tryptophan, 1.9 g; L-tyrosine, 0.3 g; L-valine, 1.3 g. The medium was prepared anaerobically under argon atmosphere, $NaHCO_3$ (2 g/liter) was added, and the mixture was reduced using 3 g/l cysteine and 1 g/l Na_2S. The final pH was 6.4. The medium was filter sterilized using a 0.22-μm-pore-size sterile filter (Millipore Filter Corp., Bedford, MA). Cultures were incubated anaerobically overnight at the optimal temperature for each: *C. saccharolyticus* DSM 8903, 70°C; *C. hydrothermalis* DSM 18901, 65°C; *C. kristjanssonii* DSM 12137, 78°C; *C. bescii* DSM 6925, 78°C; *C. kronotskyensis* DSM 18902, 70°C; *C. lactoaceticus* DSM 9545, 68°C; *C. obsidiansis* ATCC BAA-2073, 78°C. *E. coli* strain JW 261 (pDCW 68, apramycin[r]) was grown in LB broth supplemented with apramycin (50 μg/ml) with shaking at 37°C overnight. Chromosomal DNA from *C. bescii* DSM 6725 was extracted using the DNeasy® Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. The two native plasmids (pATHE01 and pATHE02) in *C. bescii* were isolated using the method described by O'Sullivan and Klaenhammer [19] with the following modifications: 200 ml of mid-log phase *C. bescii* cultures was harvested by centrifugation at 3,500*g* for 15 min and suspended in lysis buffer (containing 25% sucrose and 25 mg/ml lysozyme to enhance cell wall degradation). pDCW68 DNA was isolated from *E. coli* using a Qiagen Mini-prep Kit.

Plasmid constructions

All primers used in these constructions are listed in Table 1. pJHW006 was constructed from pSET152

Table 1 Primer list

Primer	5′–3′
JH012 forward	AGAGAGTCTAGAGGCCTTTTGCTCACATGCGTT
JH013 reverse	AGAGAGGGTACCAGGATCTCAAGAAGATCCTTTGAT
DC100 forward	TAGTCTTGATGCTTCACTGATAG
DC156 reverse	AGAGGATCCTTAAGAGATTGCTGCGTTGATA
DC165 forward	ACAGGATCCAGCTTTAATGCGGTAGTTTATCACA
DC175 forward	AGAGCTAGCTTCAACAACCAGAGACACTTGGGA
DC176 reverse	TCTGCTAGCTCCAACGTCATCTCGTTCTC
DC180 reverse	TCTGACGTCATCTTTTCCGCTGCATAACCCT
DC182 forward	AGAGACGTCAATTGAAAAAGCTTTAAAGTGTGGTGCA
DC187 reverse	CATATTGACCATCCTTTCTATGTAGA
DC188 forward	TTGAAACATTTGCTTGGGCTAAG
DC216 forward	ACAACCATGGACCAAACCGCAAAAGGAAA
DC217 reverse	TCTCCTCGAGCTCCCAACTTTCAATGTGAGAA
DC222 forward	TACAAGAAAAGCCCGTCAC
DC224 reverse	AGCTAACAATTGAGTTTACACGT
JF283 reverse	TGCAGTGTATAGCATGCAAAGCCTG
JF286 reverse	ATCCCCTTAAATTTATTTGTCTTTTAG
JF287 forward	TTTGGAAGGATGATGAACTATGAATC

(GenBank: AJ414670.1) to replace the ColEI origin of replication with the pSC101 origin. The pSC101 origin was amplified from pWSK29 (GenBank: AF016889.1) using primers JH012 with an XbaI site and primer JH013 with a KpnI site. The pSC101-containing fragment and the pSET vector digested with XbaI and KpnI were ligated to form pJHW006. Construction of pDCW68, designed for transformation of *C. bescii*, required three cloning steps as well as overlapping polymerase chain reactions (PCRs). All PCR amplifications were performed using *Pfu* Turbo DNA polymerase (Agilent Technologies). A 3.936-kb PCR product containing the pSC101 replication origin, the apramycin resistance gene, and the *oriT* (origin of transfer) was amplified from pJHW006 using primers DC176 and DC165 which contains a BamHI site. A 3.121-kb PCR product containing the *pyrBCF* region of the *C. bescii* genome was amplified from chromosomal DNA using primers DC188 and DC156 which also contained a BamHI site. The two PCR products were digested with BamHI and ligated to generate a 7.057-kb product. A 0.205-kb PCR product containing the regulatory region of a ribosomal protein (Athe 2105) was amplified using primers DC175 containing an NheI site, and DC187 using chromosomal DNA as template. A 7.066-kb fragment was amplified from the 7.057-kb product using primers DC188 and DC176 and ligated to the 0.205-kb fragment that had been digested by NheI, generating a 7.262-kb product. A 2.067-kb PCR fragment containing the 3′ flanking region of the tryptophan synthase, alpha subunit (Athe 1690) was amplified from chromosomal DNA using primers JF283 and JF287 and joined to a 2.045-kb PCR product containing the 5′

flanking region of tryptophan synthase, alpha subunit (Athe 1690) amplified from chromosomal DNA using DC182 which contained an AatII site and an overlapping primer JF286 using the high-fidelity *Pfu* DNA polymerase (Agilent Technologies). A 4.112-kb product was then generated by overlapping PCR using the two fragments and *C. bescii* genomic DNA as a template. The 7.262-kb product from the second cloning step was amplified by PCR using DC180, which contains an AatII and DC100. The 7.262-kb product and the overlapping product were digested with AatII and ligated to yield pDCW68 (11.368 kb). To construct pDCW72, the 0.981-kb CbeI (Athe 2438) open reading frame was amplified by PCR using primers DC216 and DC217 using *C. bescii* genomic DNA as template. The PCR product was digested with NcoI and Xho1 and ligated to pET24d [9], which had also been digested with NcoI and XhoI. This vector contains a His-tag sequence that is added to the C-terminus of the expressed protein. The final plasmid was sequenced to confirm that the cloned *cbeI* gene was in frame with the C-terminal His-tag followed by a translation stop codon.

Preparation of cell extracts and DNA substrates

A cell-free extracts of *C. bescii* was prepared from a 500 ml culture grown to mid-log phase, harvested by centrifugation at 6,000*g* at 4°C for 15 min and resuspended in Cel-Lytic B cell lysis reagent (Sigma) containing a protease inhibitor cocktail (Complete, EDTA-free from Roche). Extracts were sonicated on ice and then centrifuged at 13,000 rpm for 15 min at 4°C. Supernatants were removed

and used immediately for enzyme activity assays. Protein concentrations were determined using the Bio-Rad protein assay kit with bovine serum albumin (BSA) as the standard.

For in vitro methylation of DNA, pDCW68 DNA (20 μg), isolated from *E. coli* DH5α (*dam*[+], *dcm*[+]), was treated with HaeIII methyltransferase (New England Biolabs) according to the supplier's instructions. To allow complete methylation, an additional 10 units of M.HaeIII and 80 μM *S*-adenosylmethionine (SAM) was added to the reaction every 4 h of incubation at 37°C for a total of 12 h. The methyltransferase was inactivated by incubation at 65°C for 15 min. Methylated DNA was purified and concentrated using the DNA Clean & Concentrator[TM]-25 Kit (Zymo Research). The extent of protection was determined using HaeIII (New England Biolabs) according to the supplier's instructions.

The 2.56-kb fragment containing three HaeIII (5′-GG CC-3′) sites used in assays with purified CbeI was generated by PCR amplification using primers DC222 and DC224 (Table 1) from pDCW68 template. PCR products were purified and concentrated by Qiaquick PCR purification kit (Qiagen) prior to use in the restriction assays.

Endonuclease assays

Reactions were performed in 60 μl volumes at 75°C using 0.5–1.0 μg of the DNA substrate: pDCW68, methylated pDCW68, pATHE 01 or pATHE 02. For cell-free extracts, 4 μg total cell protein was incubated at 75°C in reaction buffer (10 mM Tris–HCl, pH 6.7, buffer containing 50 mM NaCl, 10 mM MgCl$_2$, 1 mM dithioerythritol, 0.01% BSA). Samples (10 μl) were withdrawn at various time points, mixed with 6× DNA gel-loading buffer (0.25% Bromophenol blue and Xylene Cyanol, 18 mM EDTA, and 30% of glycerol), and then placed at −20°C to stop the reaction. The cleavage products were separated electrophoretically on a 1.2% agarose gel.

Preparation of assay of purified CbeI protein

BL21-CodonPlus(DE3)-RILP cells (Agilent Technologies) were used for recombinant protein expression. Cells were grown at 23°C in LB broth supplemented with kanamycin (25 μg/ml) and chloramphenicol (50 μg/ml) to OD$_{260}$ 0.6 and induced by addition of 0.5 mM isopropyl *β*-D-1-thiogalactopyranoside (IPTG) at 23°C for overnight. Cells were harvested by centrifugation, resuspended in CelLytic B cell lysis reagent (Sigma) containing protease inhibitor (Complete, EDTA-free from Roche), and lysed by sonication. All purification steps were done at 4°C using the Ni–NTA Spin Kit (Qiagen) following manufacturer's instruction. Protein concentrations were determined by Bio-Rad protein assay kit as described above. Sodium

dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) and Coomassie brilliant blue G-250 staining were as described [27]. Enzyme assays with purified CbeI were carried out in 10 μl reaction volumes with NEBuffer 4 (20 mM Tris–acetate pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM dithiothreitol) and 200 ng DNA substrate. The amount of purified CbeI protein used in each reaction varied depending on the experiment and is indicated.

Bioinformatic analysis

To produce an alignment and phylogenetic tree of 46 amino acid sequences of HaeIII-like proteins, we used ClustalW, version 2 [13] which is based on the neighbor-joining (NJ) method used for phylogenetic calculations. The tree was visualized with TreeView [21]. To discover conserved motifs of groups of HaeIII-like protein sequences, we used MEME [2] and GLAM2 [8]. Default parameters were used for all analyses.

Identification of the CbeI cleavage site

DNA fragments resulting from digestion by either HaeIII or CbeI were separated electrophoretically and extracted from the gel matrix using a QIAquick gel extraction kit (Qiagen). The products were then cloned into pWSK29 [29], which had been digested with EcoRV, using the Fast-Link[TM] DNA ligation kit (Epicentre[R] Biotechnologies), and sequenced.

Results

Identification of a HaeIII-like restriction activity in *C. bescii*

A cell-free extracts from *C. bescii* was prepared [10] and incubated with pDCW68 DNA (Fig. 1a), a vector constructed for use in transformation experiments that had been isolated from *E. coli* (DH5α *dam*[+] *dcm*[+]). The plasmid DNA was completely digested within 10 min and had a restriction cleavage profile with some similarity to that of a digest with commercially available HaeIII (Fig. 1d). DNA of the two native plasmids, pATHE01 and pATHE02 (Fig. 1b, c), isolated from *C. bescii* were not digested when incubated with the same cell-free extract nor were they digested by commercially available HaeIII endonuclease (Fig. 1e), suggesting the presence of HaeIII methyltransferase-like activity in *C. bescii*. In addition, pDCW68 DNA that had been methylated in vitro by commercially available HaeIII methyltransferase was protected from cleavage by either the commercially available HaeIII
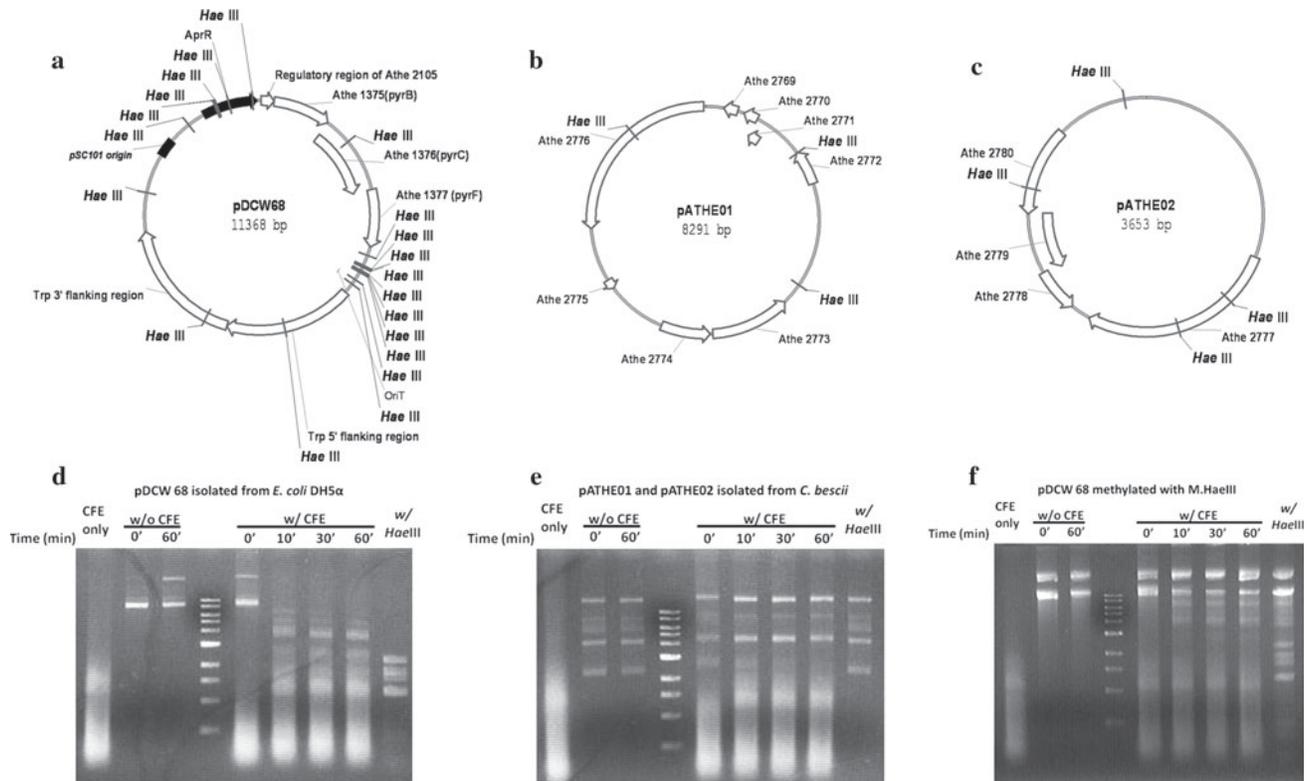
Fig. 1 Detection of a HaeIII-like restriction/modification system in *C. bescii*. **a** pDCW68; **b** pATHE01; **c** pATHE02; **d** pDCW68 isolated from *E. coli* DH5α incubated with CFE from *C. bescii*; **e** pATHE01 and pATHE02 isolated from *C. bescii* incubated with CFE from *C. bescii*; **f** pDCW68 isolated from *E. coli* treated with M.HaeIII and incubated with CFE from *C. bescii*. Features originating from *E. coli* are shaded in *black*, and features from *C. bescii* in white; Apr^R, apramycin resistant gene cassette; Trp, Athe 2105-tryptophan synthetase α subunit; *OriT*, origin of transfer for conjugation; pSC101, low copy replication origin in *E. coli*; HaeIII restriction sites are indicated. All incubation times were as indicated

restriction enzyme or cell-free extracts from *C. bescii* (Fig. 1f), implying the presence of cognate methyltransferase activity in *C. bescii*. In support of the notion that *C. bescii* contains a HaeIII-like enzyme is the fact that HaeIII recognition sites (GGCC) are much less abundant in the *C. bescii* genome sequence than would be expected based on random nucleotide composition (0.468 observed/expected). Organisms that produce restriction enzymes often have a bias against the presence of the recognition sequences of those enzymes in their genomes [17, 26]. We have named this HaeIII-like restriction endonuclease from *C. bescii* as CbeI.

Cloning, expression, and purification of CbeI from *C. bescii*

A query of the *C. bescii* genome using the GenBank [3] database as well as REBASE [25] identified a methyltransferase (Athe 2437) adjacent to a candidate gene for a HaeIII-like restriction endonuclease (Athe 2438) (Fig. 2a),

which is a common feature of type II restriction/modification gene arrangements [12].

The open reading frame encoding CbeI (Athe 2438) was cloned into an *E. coli* expression vector, pDCW72 (Fig. 2b), under transcriptional control of the T7 promoter to allow regulated expression of the restriction endonuclease, as it would be expected to be toxic to cells that did not contain the corresponding methyltransferase. In fact, initial attempts to clone the gene encoding a His-tagged version, CbeI-His$_6$, at 37°C, even without induction of the T7 promoter, failed. Because *C. bescii* grows optimally near 80°C, we investigated the possibility that cloning and expression of CbeI at lower temperatures would be more efficient, because at lower temperatures the enzyme might be inactive or nonfunctional. When experiments were done at 23°C (the equivalent of expressing *E. coli* enzymes at −6°C), both cloning and expression were successful. Since there are also significant differences in codon usage between *C. bescii* (35.2% GC content) and *E. coli* (50.5% GC content), BL21-CodonPlus(DE3)-RIPL cells which contain rare transfer RNAs (tRNAs) were used for

Fig. 2 Cloning, expression, and purification of CbeI. **a** The region surrounding the location of *cbeI* in the *C. bescii* genome. **b** pDCW72; KanR, kanamycin resistance gene; bacteriophage T7 promoter and T7 terminator; *lacI*, the gene for the lactose repressor protein; ColEI, origin of replication derived from pBR322 **c** *Lane 1* protein molecular weight standards; *lane 2* 15 ng of purified CbeI protein displayed on a 10–20% Tris–HCl gradient gel (Criterion[TM] Precast Gel; Bio-Rad Laboratories, Hercules, CA)
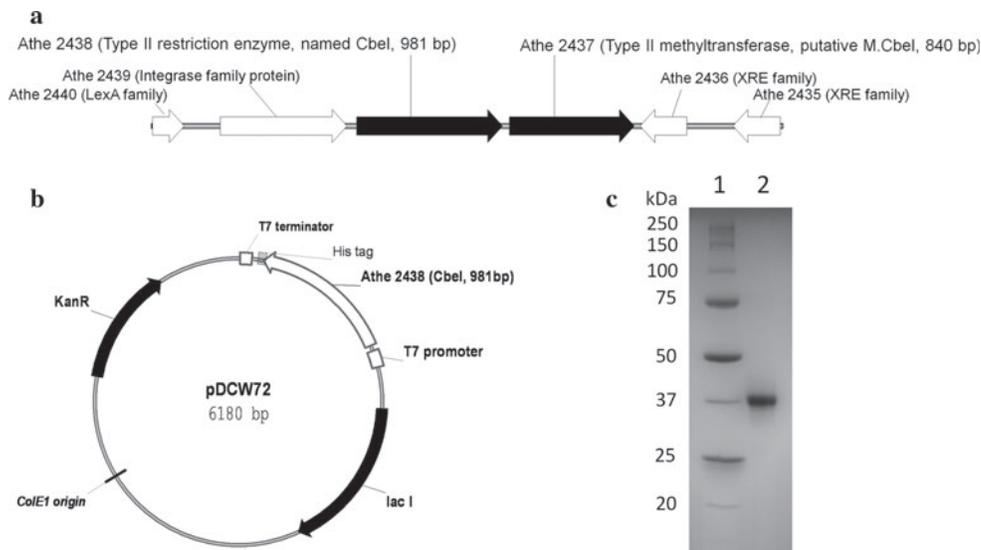

expression. Preparations of purified C-terminal His-tagged CbeI contained a single band on an SDS–PAGE gel (Fig. 2c) with molecular mass of approximately 38 kDa, which is consistent with the calculated value of 37.9 kDa determined using the Compute pI/Mw analysis tool.

Functional analysis and temperature optimum of the purified CbeI endonuclease activity

A DNA substrate containing three HaeIII restriction sites (Fig. 3a), which should generate fragments of 1,393, 772, 293, and 106 bp when digested, was used in restriction digestion assays. Purified CbeI protein (12–48 ng) was incubated with 50 mM potassium acetate, 20 mM Tris–acetate, 10 mM magnesium acetate, 1 mM dithiothreitol at pH 7.9, and 100 ng DNA substrate in 10 µl (Fig. 3a), and incubated for 10–20 min at 75°C. These conditions were chosen based on the reaction conditions of other isoschizomers of HaeIII from thermophiles [18, 22] and on the composition of NEBuffer 4 that is used with commercially available HaeIII restriction endonuclease.

Experiments with cell-free extracts indicated that CbeI cleavage activity occurred within 10 min even with low protein concentrations (4 µg total cell protein in 60 µl). Purified protein showed a similar time course, and there was no difference in activity between 12 and 48 ng (in 10 µl) of purified protein in the enzyme assays. The experiment shown in Fig. 3b used 24 ng (in 10 µl) of purified protein. In assays with less than 2.5 ng (in 10 µl) of protein, no activity was detected.

During the course of cloning and expressing CbeI in *E. coli*, it was noted that the gene was not toxic to cells grown at 23°C but it was at 37°C, suggesting that the enzyme was less active at low temperature. To determine the optimal temperature for CbeI activity, purified protein
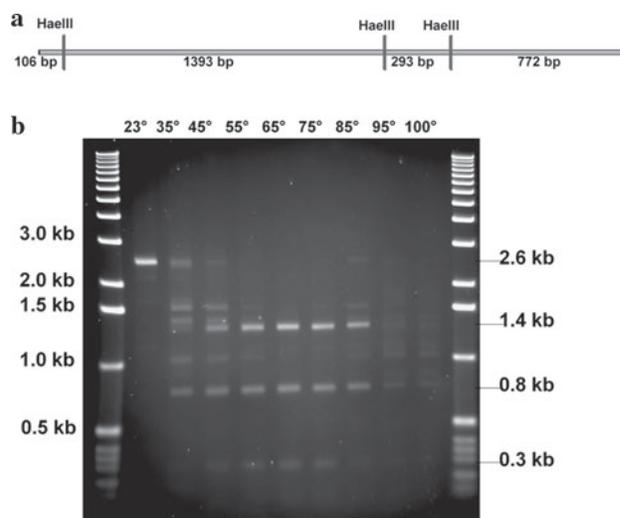
Fig. 3 Temperature profile of purified CbeI endonuclease activity. **a** The 2.558-kb DNA substrate synthesized as described in the "Materials and methods." HaeIII cleavage sites are marked by *vertical lines*, and predicted cleavage fragment sizes are indicated below the *line*. **b** pDCW72 used for His-tagged expression of CbeI. The substrate was incubated for 10 min with 0.5 ng protein at the temperatures indicated, and the cleavage products were separated on 1.2% agarose gel. The position of the full-length undigested fragment and the three major cleavage products derived by digestion with CbeI are indicated by *arrows*

was incubated with the DNA substrate at temperatures ranging from 23°C to 100°C. As shown in Fig. 3b, the enzyme shows the highest activity between 75°C and 85°C, exhibiting partial digestion activity at 45°C or below and nonspecific activity above 85°C. As predicted, no activity was detected at 23°C. The optimum temperature for commercially available HaeIII activity is 37°C, and the enzyme is inactivated by treatment at 80°C for 20 min. In contrast, CbeI is optimally active at 75–85°C and is stable for more

than 30 min at 75°C (Fig. 3b). CbeI did not lose activity after storage for more than 1 week at up to 35°C and was inactivated by incubation at 100°C for 5 min.

CbeI is an isoschizomer of HaeIII

HaeIII recognizes a sequence that includes GGCC and cleaves between the G and C, leaving a blunt-ended fragment. To determine the cleavage site of CbeI, a DNA fragment containing three HaeIII recognition/cleavage sites (Fig. 3a) was used as substrate with purified CbeI enzyme. The cleavage products were either ligated directly to pWSK29 that had been digested with EcoRV leaving a blunt end, or first treated with the Klenow fragment of DNA polymerase prior to ligation. The sequence of the region containing the cloning site revealed the dinucleotide 5′-CC-3′ adjacent to the EcoRV site and the dinucleotide 5′-GG-3′ adjacent to the EcoRV site, identifying 5′-GG/CC-3′ as the CbeI cleavage site. The same result was obtained in eight independent cloning experiments, and we conclude that CbeI is an isoschizomer of HaeIII.

Cluster analysis of HaeIII-like proteins reveals that CbeI is a member of a new subfamily of HaeIII-like enzymes

A search of the REBASE and NCBI databases revealed 231 HaeIII-like proteins. Of those, 183 had been fully or partially characterized and 48 were putative isoschizomers predicted from their DNA sequences. Of the 57 proteins for which there was sequence information available, subfamilies of the same genus and species were removed, leaving 46 sequences used in this analysis. A phylogram based on protein sequence alignments is shown in Fig. 4. CbeI falls into a distinct group of 13 proteins, 5 of which had previously been grouped (pfam09556) based on overall sequence similarity using the Conserved Domain Database [15]. Our own motif-based sequence analysis (MEME Suite and GLAM2) of this group of 13 proteins with other HaeIII-like proteins identified three conserved motifs shared by 11 of the 13 (Fig. 5) not present in other HaeIII-like proteins. One protein (Bgr_19490) had motifs 1 and 2, and one (GOS_4010239) had only motif 1. Attempts to identify structural motifs of CbeI based on three-dimensional (3D) homology modeling using 3D Jigsaw, CPHmodels, swiss-model, and ESy-Pred3D failed to find significant matches with any known structural motifs. As with other members of this diverse group of proteins, attempts to assign CbeI to a super-family of endonucleases to identify functional domains and conserved active-site amino acids were also unsuccessful.

Evidence for HaeIII-like restriction/modifications systems in other *Caldicellulosiruptor* species

To investigate whether other *Caldicellulosiruptor species* contained HaeIII-like restriction/modification activities, total DNA was isolated from a number of *Caldicellulosiruptor* species and incubated with commercially available HaeIII restriction endonuclease. As shown in Fig. 6, *C. saccharolyticus* DSM 8903, *C. hydrothermalis* DSM 18901, *C. kristjanssonii* DSM 12137, and *C. bescii* DSM 6925 were resistant to HaeIII nuclease, while *C. kronotskyensis* DSM 18902, *C. lactoaceticus* DSM 9545, and *C. obsidiansis* ATCC BAA-2073 were sensitive. These preliminary data suggest that this HaeIII-like restriction/modification system may be widespread among members of this genus.

Discussion

In a screen of *C. bescii* cell-free extracts for restriction endonuclease activities we discovered a potent HaeIII-like restriction activity with novel features. Plasmid DNA from *E. coli*, but not plasmid DNA from *C. bescii*, was digested within 10 min of incubation with these extracts, and treatment with HaeIII methyltransferase protected the DNA from cleavage, suggesting the existence of a HaeIII-like restriction-modification system in *C. bescii*. Bioinformatic analysis of the *C. bescii* genome identified a gene encoding a protein homologous to the HaeIII endonuclease and an adjacent gene encoding a type II DNA methyltransferase. The gene for the endonuclease was cloned and expressed in *E. coli* with a His-tag. Purified enzyme from *E. coli* was optimally active between 55°C and 85°C and was stable at 35°C for more than 1 week. The cleavage site of the enzyme was determined to be GG/CC, suggesting that it is an isoschizomer of HaeIII, and we have named this enzyme CbeI. A phylogram of CbeI with other HaeIII-like enzymes identified a new subfamily of these enzymes with unique features.

The cloning and expression of CbeI in *E. coli* presented some challenges, as does the expression of other toxic genes including other restriction endonucleases [12, 24]. In addition to using BL21-CodonPlus(DE3)-RIPL cells to compensate for differences in codon usage between *E. coli* and *C. bescii*, we took advantage of the fact that CbeI is from an extreme thermophile and would be expected to have minimal activity at temperatures significantly below the growth temperature of *C. bescii* ($T_{opt} \approx 80$°C). This appeared to be the case, since expression of CbeI was apparently toxic to *E. coli* cells grown at 37°C but was expressed efficiently at 23°C. This strategy may be useful for expressing toxic genes derived from thermophilic
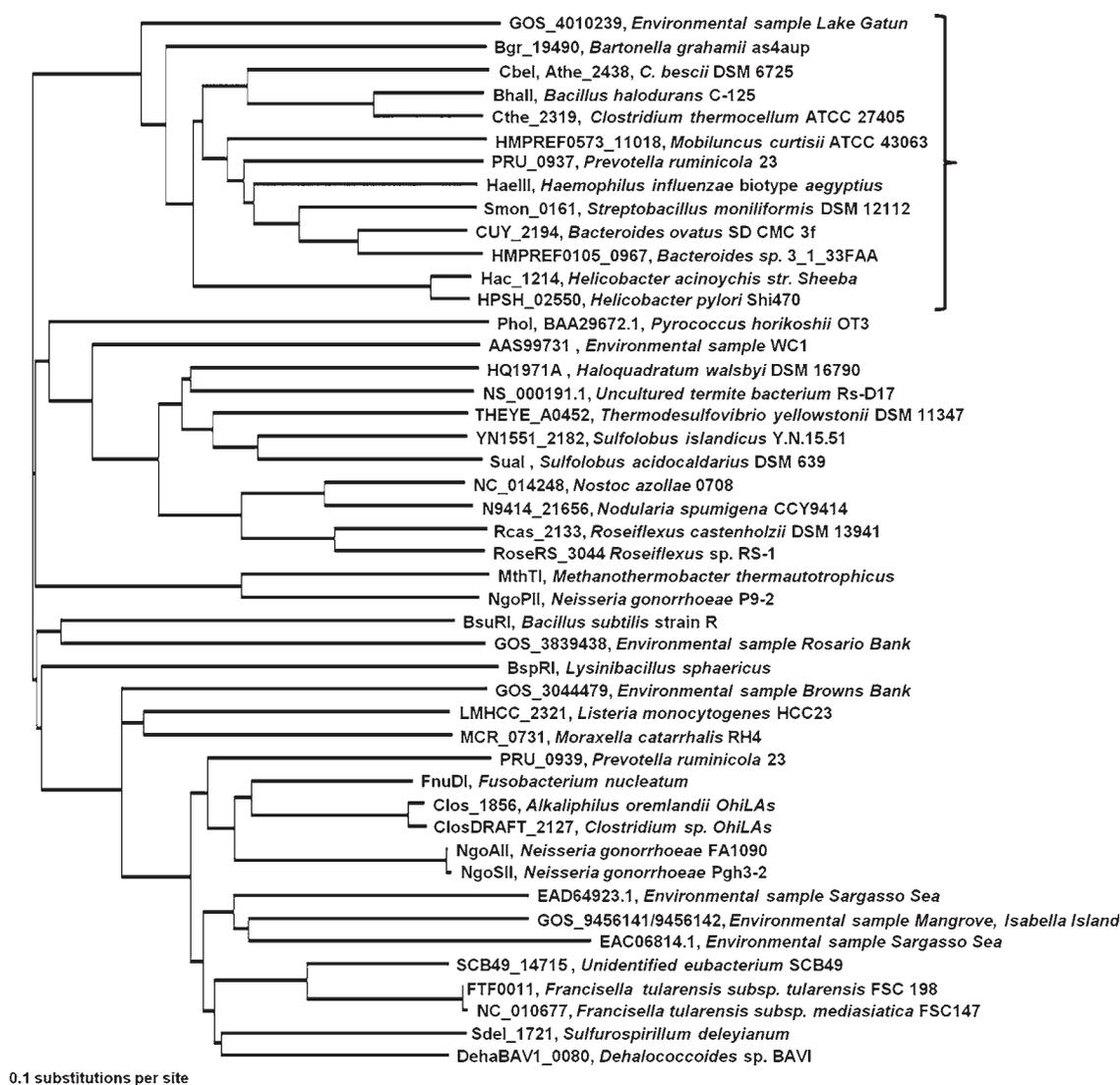
Fig. 4 Phylogram alignment of 46 HaeIII-like restriction enzymes. The host organism for each restriction enzyme is indicated as well as the protein name, when available; otherwise, the GenBank locus tag or accession number is given. The distance scale is indicated by a *bar* defining the distance for 0.1 amino acid substitution per site. The *bracketed* organisms represent those containing this new subfamily of HaeIII-like enzymes that includes CbeI

organisms in *E. coli*, eliminating the need for complicated highly regulated expression systems.

Since the first description of the HaeIII restriction enzyme in 1972 [16], more than 200 isoschizomers have been reported or predicted [25]. Of these, fewer than 40 are from thermophiles (organisms that have $T_{opt} \geq 50°C$), and only three of these have been characterized: MthTI [18] from *Methanobacterium thermoformicicum* THF (T_{opt} 55°C), NspLKI [32] from *Nocardia* species LK (T_{opt} 50°C), and SuaI [22] from *Sulfolobus acidocaldarius* (T_{opt} 82°C). A fourth, PhoI from *Pyrococcus horikoshii* (T_{opt} 98°C), is commercially available (New England Biolabs), but there are no reports on this enzyme in the literature. Unlike HaeIII itself, which is optimally active at

37°C and is inactivated by heating to 80°C, CbeI was optimally active in the range 75–85°C and required incubation at 100°C for 5 min for inactivation. The fact that CbeI isolated from *E. coli* is thermostable suggests that this feature is due to its conformation, hydrophobic, electrostatic, or other properties rather than by association with other proteins or cofactors in *C. bescii*.

HaeIII-like enzymes are widespread in both the archaea and bacteria. A previous study showed that the genes encoding NgoPII from *Neisseria gonorrhoeae*, a bacterium, and MthTI from *Methanobacterium thermoformicicum*, an archaeon, have unexpectedly high similarity (54.5% nucleotide identity), suggesting horizontal gene transfer [18]. In fact, a phylogenetic tree based on protein sequence
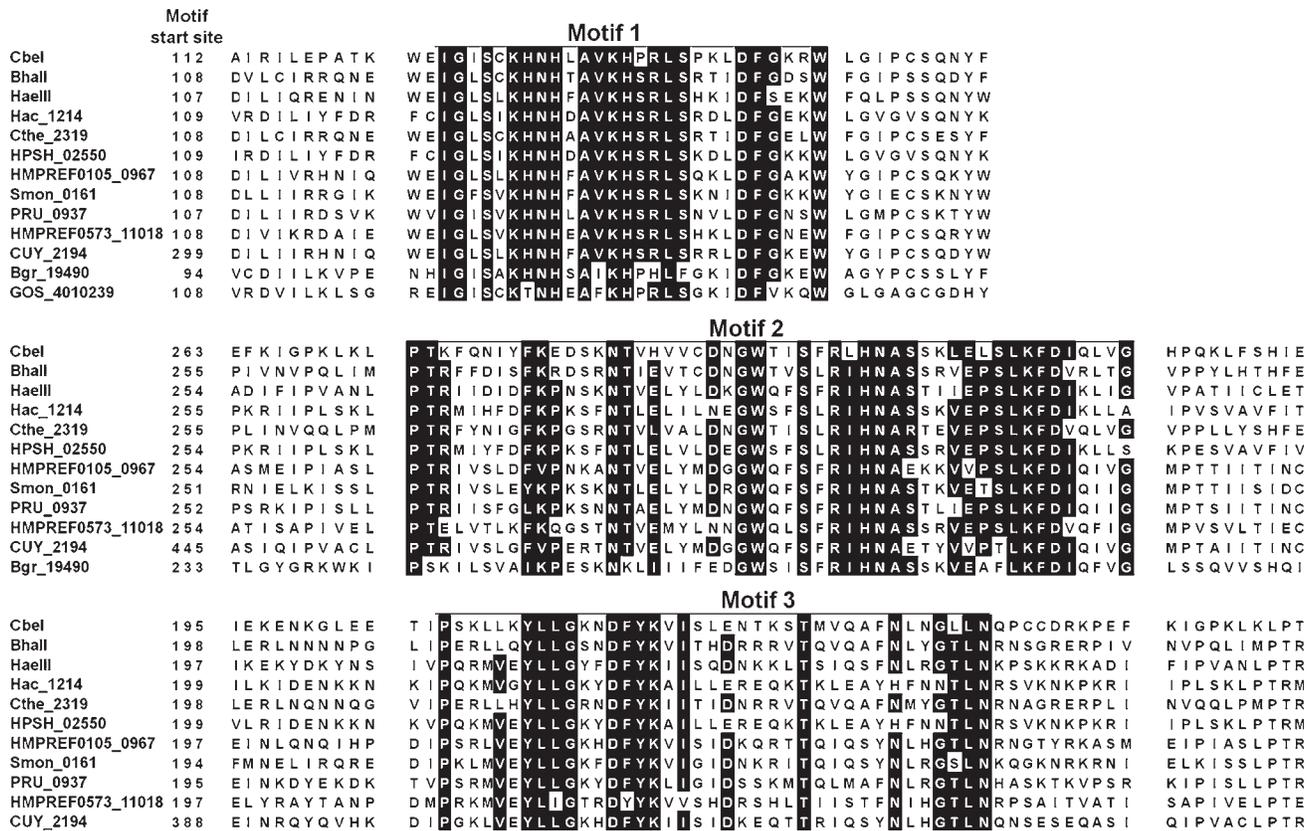
Fig. 5 Amino acid sequence alignment of CbeI with the subgroup identified in Fig. 4. Amino acid identity is shown as shaded areas with the position of the motif within the protein sequence. Motif start site is indicated. Motif 1: 7.0e-220, motif 2 1.9e-278, motif 3 1.6e-185
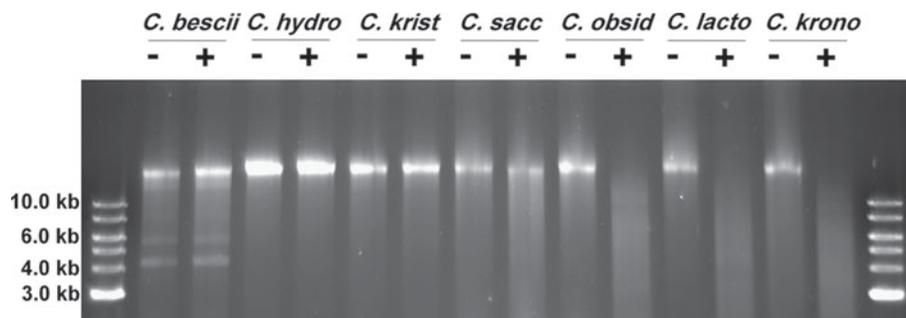
Fig. 6 Distribution of HaeIII-like restriction/modification systems in *Caldicellulosiruptor species*. Total DNA isolated from seven different species were incubated (−) without or (+) with commercially available HaeIII endonuclease at 37°C for 1 h according to the manufacturer's instructions (NEB). *C. bescii*; *C. hydro*, *C. hydrothermalis*; *C. krist*, *C. kristjansonii*; *C. sacc*, *C. saccharolyticus*; *C. obsid*, *C. obsidiansis*; *C. lacto*, *C. lactoaceticus*; *C. krono*, *C. kronotskyensis*. Also visible in the *C. bescii* lanes are the undigested native plasmids from that strain, pATHE01 (8.3 kb) and pATHE02 (3.6 kb)

similarity of HaeIII-like proteins (Fig. 4) identified a subgroup that includes four proteins from archaea (*Pyrococcus horikoshii* OT3, *Sulfolobus islandicus*, *Sulfolobus acidocaldarius*, and *Methanothermobacter thermautotrophicum*) and nine from bacteria, suggesting that there may have been cross-domain horizontal gene transfer for these proteins. In support of this notion is the fact that the GC content of some of the genes encoding HaeIII-like proteins is significantly different from that of their host organism

chromosomes: *Mobiluncus curtisii* (55% genome, 46% HMPREF0573), *Prevotella ruminicola* (47% genome, 32% PRU_0939), and *Roseiflexus castenholzii* (60% genome, 52% Rcas_2133). HaeIII-like enzymes are also widespread in both archaeal and bacterial thermophiles, such as *Clostridium thermocellum* ATCC 27405, *Methanothermobacter thermautotrophicum* THF [18], *Nocardia species* LK [32], *Roseiflexus castenholzii* DSM 13941, *Sulfolobus islandicus*, *Sulfolobus acidocaldarius* DSM 639 [22], *Pyrococcus*

horikoshii OT3, and *Thermodesulfovibrio yellowstonii* DSM 11347.

In an analysis of 46 HaeIII-like proteins, those most similar to CbeI were found in other bacteria, both Gram-positive and Gram-negative. CbeI exhibits the highest amino acid sequence similarity ($\sim 60\%$) with the HaeIII-like proteins from *Bacillus halodurans* (BhaII) and *Clostridium thermocellum* (Cthe_2319). Examination of genomic DNA isolated from seven different *Caldicellulosiruptor* species showed that four of the seven were resistant to HaeIII cleavage, indicating that HaeIII-like restriction-modification systems may be widespread in members of this genus.

An amino acid sequence alignment of CbeI and the 12 closely related HaeIII-like proteins (bracketed in Fig. 4) revealed highly conserved residues that define three previously unrecognized motifs that may play a role in their structure or catalytic activity. Although these 13 HaeIII-like proteins could not be reliably matched to any other known protein structure or to the five known type II restriction endonuclease superfamilies [PD-(D/E)XK, HNH, PLD, GIY_YIG, and HALFPIPE] [20], these observations make CbeI an interesting candidate for structural analyses, since it may possess a novel tertiary structure. This new subgroup identified in our analysis that includes CbeI defines a new subfamily of structurally or functionally related proteins in this diverse group of enzymes. The results presented herein also have important implications for the development of methods of genetic transformation for this interesting and biotechnologically important group of relatively uncharacterized organisms.

References

1. Accetto T, Peterka M, Avgustin G (2005) Type II restriction modification systems of *Prevotella bryantii* TC1–1 and *Prevotella ruminicola* 23 strains and their effect on the efficiency of DNA introduction via electroporation. FEMS Microbiol Lett 247:177–183

2. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28–36

3. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2009) GenBank. Nucleic Acids Res 37:D26–D31

4. Bron S, Murray K (1975) Restriction and modification in *B. subtilis*. Nucleotide sequence recognised by restriction endonuclease R. Bsu R from strain R. Mol Gen Genet 143:25–33

5. Chen CK, Boucle CM, Blaschek HP (1996) Factors involved in the transformation of previously non-transformable *Clostridium perfringens* type B. FEMS Microbiol Lett 140:185–191

6. Cue D, Lam H, Dillingham RL, Hanson RS, Flickinger MC (1997) Genetic manipulation of *Bacillus methanolicus,* a gram-positive, thermotolerant methylotroph. Appl Environ Microbiol 63:1406–1420

7. Donahue JP, Israel DA, Peek RM, Blaser MJ, Miller GG (2000) Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. Mol Microbiol 37:1066–1074

8. Frith MC, Saunders NF, Kobe B, Bailey TL (2008) Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput Biol 4:e1000071

9. Hethke C, Geerling AC, Hausner W, de Vos WM, Thomm M (1996) A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res 24:2369–2376

10. Jennert KC, Tardif C, Young DI, Young M (2000) Gene transfer to *Clostridium cellulolyticum* ATCC 35319. Microbiology 146(Pt 12):3071–3080

11. Kataeva IA, Yang SJ, Dam P, Poole FL 2nd, Yin Y, Zhou F, Chou WC, Xu Y, Goodwin L, Sims DR, Detter JC, Hauser LJ, Westpheling J, Adams MW (2009) Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. J Bacteriol 191:3760–3761

12. Kong H, Lin LF, Porter N, Stickel S, Byrd D, Posfai J, Roberts RJ (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. Nucleic Acids Res 28:3216–3223

13. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

14. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577

15. Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH (2009) CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 37:D205–D210

16. Middleton JH, Edgell MH, Hutchison CA 3rd (1972) Specific fragments of phi X174 deoxyribonucleic acid produced by a restriction enzyme from *Haemophilus aegyptius*, endonuclease Z. J Virol 10:42–50

17. Nobusato A, Uchiyama I, Kobayashi I (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. Gene 259:89–98

18. Nölling J, de Vos WM (1992) Characterization of the archaeal, plasmid-encoded type II restriction-modification system MthTI from *Methanobacterium thermoformicicum* THF: homology to the bacterial NgoPII system from *Neisseria gonorrhoeae*. J Bacteriol 174:5719–5726

19. O'Sullivan DJ, Klaenhammer TR (1993) Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactobacillus* spp. Appl Environ Microbiol 59:2730–2733

20. Orlowski J, Bujnicki JM (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. Nucleic Acids Res 36:3552–3569

21. Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

22. Prangishvili DA, Vashakidze RP, Chelidze MG, Gabriadze I (1985) A restriction endonuclease SuaI from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett 192:57–60

23. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol Microbiol 46:439–452

24. Rasko T, Der A, Klement E, Slaska-Kiss K, Posfai E, Medzihradszky KF, Marshak DR, Roberts RJ, Kiss A BspRI restriction endonuclease: cloning, expression in *Escherichia coli* and sequential cleavage mechanism. Nucleic Acids Res

25. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 38:D234–D236

26. Rocha EP, Danchin A, Viari A (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. Genome Res 11:946–958

27. Sedmak JJ, Grossberg SE (1977) A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G250. Anal Biochem 79:544–552

28. Svetlichnyi VA, Svetlichnaya TP, Chernykh NA, Zavarzin GA (1990) *Anaerocellum thermophilum*, new genus new species, an extremely thermophilic cellulolytic eubacterium isolated from hot springs in the valley of geysers (Russian SFSR, USSR). Mikrobiologiya 59:598–604

29. Wang RF, Kushner SR (1991) Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene 100:195–199

30. Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MW (2009) Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. Appl Environ Microbiol 75:4762–4769

31. Yang SJ, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MW (2009) Reclassification of '*Anaerocellum thermophilum*' as *Caldicellulosiruptor bescii* strain DSM 6725T sp. nov, Int J Syst Evol Microbiol

32. Zabaznaya EV, Zheleznaya LA, Svad'bina IV, Matvienko NI (1999) Site-specific endonuclease NspLKI is an isoschizomer of endonuclease HaeIII. Biochemistry (Mosc) 64:189–193