J Ind Microbiol Biotechnol
DOI 10.1007/s10295-013-1345-8

SIMB
Society for Industrial Microbiology and Biotechnology

Isolation and bioinformatic analysis of a novel transposable element, IS*Cbe*4, from the hyperthermophilic bacterium, *Caldicellulosiruptor bescii*

Minseok Cha · Hao Wang · Daehwan Chung · Jeffrey L. Bennetzen · Janet Westpheling

Received: 2 August 2013 / Accepted: 9 September 2013

Abstract *Caldicellulosiruptor bescii* is an anaerobic thermophilic bacterium of special interest for use in the consolidated bioprocessing of plant biomass to biofuels. In the course of experiments to engineer pyruvate metabolism in *C. bescii*, we isolated a mutant of *C. bescii* that contained an insertion in the L-lactate dehydrogenase gene (*ldh*). PCR amplification and sequencing of the *ldh* gene from this mutant revealed a 1,609-bp insertion that contained a single open reading frame of 479 amino acids (1,440 bp) annotated as a hypothetical protein with unknown function. The ORF is flanked by an 8-base direct repeat sequence. Bioinformatic analysis indicated that this ORF is part of a novel transposable element, IS*Cbe*4, which is only intact in the genus *Caldicellulosiruptor*, but has ancient relatives that are present in degraded (and previously unrecognized) forms across many bacterial and archaeal clades.

Keywords Transposable elements · IS*Cbe4* · *Caldicellulosiruptor bescii* · Lactate dehydrogenase · Bioinformatics

M. Cha and H. Wang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10295-013-1345-8) contains supplementary material, which is available to authorized users.

M. Cha · H. Wang · D. Chung · J. L. Bennetzen · J. Westpheling (✉)
Department of Genetics, University of Georgia, Athens, GA 30602, USA
e-mail: janwest@uga.edu

M. Cha · H. Wang · D. Chung · J. L. Bennetzen · J. Westpheling
The BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Introduction

Transposable elements are virtually ubiquitous in bacterial genomes. They consist of insertion sequences (ISs) that range in size from 600 to 3,000 bp and have been grouped into 20 major families. These elements can be responsible for dramatic changes in genes and genomes by their movement or their use as ectopic recombination sites [18, 23, 28]. They are also involved in shaping and reshuffling genomes by facilitating horizontal gene transfer [23]. IS elements typically contain a gene for a transposase with inverted repeat (IRs) sequences positioned at both ends of the IS. These sequences are recognized by the transposase which cleavages the site catalyzing the movement of the transposable element (TE) by a cut-and-paste mechanism [1, 18], usually generating a target site duplication (TSDs) at the site of insertion [1].

IS elements also have many important applied uses. They may be used for insertional mutagenesis to identify the genes involved in virtually any scorable process. For instance, their use for insertion mutagenesis has been employed to identify virulence functions in many pathogens, including *Escherichia coli* [9], *Vibrio cholerae* [25], *Yersinia pestis* [12], and *Clostridium perfringens* [4]. Their relatively high level of instability has made them useful as markers for restriction fragment length polymorphism studies, including genetic mapping and relatedness characterizations [15, 17, 24].

IS finder (http://www-is.biotoul.fr) currently shows that over 1,500 different ISs have been identified and are distributed in almost all prokaryotes, including over 295 eubacterial and archaeal species [22]. Among the thermophilic bacterial species, many of these ISs have been found in *Clostridium thermocellum* [15]. However, only a few have been described in other *Clostridial* species, including

Published online: 01 October 2013

four in *Clostridium perfringens* [4], one in *Clostridium beijerinckii* NCIMB 8052 [16], and one in *Clostridium thermocellum* [31]. Other thermophilic bacterial species, such as *Thermoanaerobacter tengcongensis* and *Thermotoga maritime*, contain elements annotated as ISs but none have been shown to be active. We recently reported the discovery of the first active element in *Caldicellulosiruptor* species, *C. hydrothermalis* [6], and here we report the isolation and identification of an active and novel IS element in *C. bescii*. The element, previously annotated as a hypothetical protein of unknown function, was discovered as an insertion in the L-lactate dehydrogenase gene (*ldh*) of *C. bescii*.

Materials and methods

Strains, growth conditions, and molecular techniques

Caldicellulosiruptor bescii strains were grown in modified DSMZ 516 medium or LOD (low osmolality defined growth medium) [11], at pH 7.0. Liquid cultures for genomic DNA extraction were grown from a 0.5 % inoculum or a single colony into anaerobic culture bottles degassed with five cycles of vacuum and argon incubated at 75 °C. Genomic DNA from *C. bescii* was purified using the *Quick-gDNA* MinPrep Kit (Zymo Research Corporation, Irvine, CA, USA). Strains used in this study were JWCB001 (*C. bescii* wild-type DSM 6725) [26], JWCB005, (DSM 6725 *ΔpyrFA*) [8], and JWCB018 (DSM 6725 *ΔpyrFA ΔcbeI*::IS*Cbe*4) [7].

Isolation and identification of a transposable element from *C. bescii*

PCR amplification of the *C. bescii ldh* locus used primers DC352 (5′-TCAACATAGAACCTCCCCA-3′) and DC355 (5′-TTGCAACAGCTCAAAGTAGCA-3′). All PCR amplifications were performed using *Pfu* Turbo DNA polymerase (Agilent Tech., Santa Clara, CA, USA). The PCR products were sent to Macrogen (Macrogen Corp., Rockville, MD, USA) for sequencing. The annotated genomic DNA sequences of the eight sequenced *Caldicellulosiruptor* spp. were downloaded from EMBL-EBI (http://www.ebi.ac.uk/). The DNA sequence of the insertion in the *ldh* locus of *C. bescii* was used to query the database of all available microbial genomes (total number 2,503 = 2,428 Bacteria + 122 Archaea + 426 Eukaryotes) using the NCBI blast server (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Homologous elements of the sequence were discovered by BLASTN (E value = 10^{-5}) analysis in four of the eight *Caldicellulosiruptor* genomes but nowhere else in the database. Identities of the sequences were aligned by MUSCLE [10] using default parameters and manual inspection to detected potential TIR (terminal inverted repeats) and TSD (target site duplications).

Phylogenetic analysis and orthologous group construction

Multiple sequence alignments of DNA sequences of the elements were constructed by MUSCLE [10] using default parameters and poorly aligned regions were removed by trimAL [5] with the option “automated1”. MEGA5 [27] was then used to construct the un-rooted neighbor-joining tree [22] (Model: Kimura 2-parameter model [14]; rate and pattern: uniform rate and homogeneous pattern among lineages; Gap/missing data treatment: pair-wise deletion). Orthologous gene groups of *C. bescii* and *C. kronotskyensis*, *C. bescii* and *C. owensensis*, and *C. bescii* and *C. lactoaceticus* were built by Inparanoid 4.1 [20]. Only one-to-one orthologues (defined as the best hit in both directions) were kept for subsequent sequence identity analysis.

Consensus sequences and global alignments

Consensus sequences of the four subfamilies were constructed using the cons program in the EMBOSS package [21]. The identity values between consensus sequences and regular orthologous genes were estimated by the Needleman–Wunsch algorithm [19] using the needle program in the EMBOSS package [21].

Results and discussion

Isolation and identification of a transposable element from *C. bescii*

During the course of constructing deletions by marker replacement in *C. bescii* [7], we used PCR amplification to check mutant constructions. Amplification of the *ldh* locus in one of these mutants, JWCB018, revealed an insertion of 1,609 bp in the middle of the *ldh* gene (Fig. 1a). This insertion contained a single open reading frame of 476 amino acids (Fig. 1b, c and S1), previously annotated as a hypothetical protein (Cbes_0092) of unknown function [13]. A total of 68 intact homologous elements and 13 truncated elements were found in the genomes of *C. hydrothermalis* (1 intact, 1 truncated), *C. kristjanssonii* (1 intact, 0 truncated), *C. saccharolyticus* (1 intact, 4 truncated), *C. kronotshyensis* (6 intact, 2 truncated), *C. owensensis* (27 intact, 2 truncated), *C. lactoaceticus* (23 intact, 1 truncated) and *C. bescii* (9 intact, 3 truncated) and were grouped into two families and four sub-families. Analysis of the ORF and the sequence of the insertion itself revealed that it is actually an active insertion sequence apparently unique to this genus and that the ORF encodes a transposase, also unique to

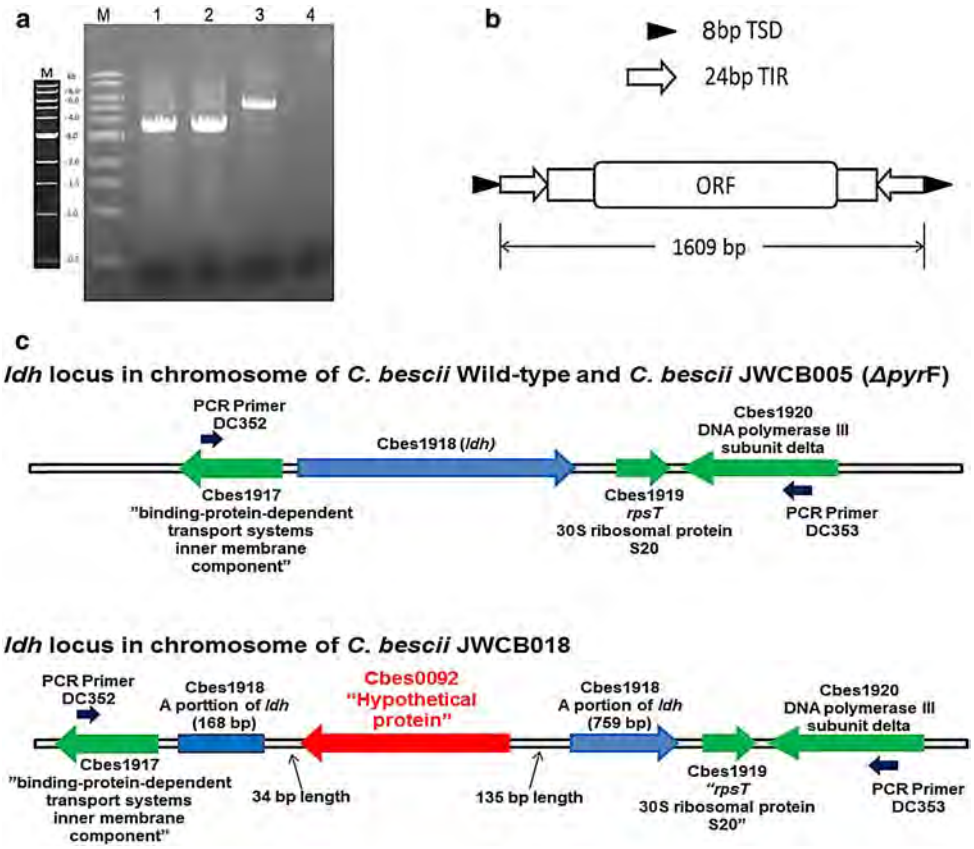

Fig. 1 Confirmation of IS element insertion within *ldh* (Cbes1819) open reading frame in *Caldicellulosiruptor bescii* and structure of IS*Cbe*4. **a** Gel depicting PCR products of the *ldh* locus for an IS element insertion and excision mutant compared to wild type and the parent strain using primers DC352 and DC353. *M*, 1 kb DNA Ladder; *Lane 1*, *ldh* locus amplification of wild type (3.3 kb); *lane 2*, *ldh* locus amplification of JWCB005 (*pyr*FA deletion strain, 3.3 kb); *lane 3*, *ldh* locus amplification of IS element insertion strain, JWCB018 (5.0 kb); *lane 4*, no template for PCR native control. **b** Structure of IS*Cbe*4 in *ldh*. **c** Diagrams of *ldh* loci in chromosomes of *Caldicellulosiruptor bescii* strains

this genus. We have named this element IS*Cbe*4. IS*Cbe*4 is 1,609 bp in length and contains terminal inverted repeats of (TIRs) of 24 bp and a target site duplication (TSD) of 8 bp (Fig. 1b, c). The identity between the two TIRs is 71 %. We note that these IRs are not identical at the two most terminal bases, making its movement something that would not be predicted from sequence analysis. The element contains an ORF of 1,440 bp (479 aa), hereafter called IS*Cbe4*_ORF. A genome-wide survey reveals that the standard size of the TSD of this element is 9 bp (Table S1).

Elucidation of the structure of IS*Cbe4* and taxonomic distribution of IS*Cbe4*_ORF

The structure of IS*Cbe*4 was similar to eukaryotic TIR DNA transposons. IS*Cbe4*_ORF did not show conservation with any known transposase (blastp, *E* value = 10^{-5} against IS FINDER [23] and the in-house TE domain database). IS*Cbe4*_ORF had 98 significant (tblastn, *E* value = 10^{-5}) hits in the NCBI microbial genome database. The 98 sequences were from 42 genera that covered a wide spectrum of bacteria, mainly in Firmicutes. Thermophilic organisms accounted for ~60 % of all hits. The taxonomic information of the 42 genera is represented in Table S3. The conserved protein domain UPF0236 (pfam06782) was identified in the ORF. This domain is found in 480 bacterial and two archaeal species but the function of the domain is unknown.

Although IS*Cbe*4-like proteins were widely distributed in bacteria and some Archaea, nucleotide-level comparisons to all sequenced microbial genomes showed IS*Cbe4* and related elements were only found in the genus *Caldicellulosiruptor*. Blastn (hit size > 100 bp; *E* value < 10^{-10}) against the eight sequenced *Caldicellulosiruptor* genomes identified 53 elements that had IS*Cbe*4 TIRs and TSD, 15 had TIR but no TSD and 13 were truncated versions of the element (Table S1). The 81 elements were distributed in seven of the eight sequenced *Caldicellulosiruptor* genomes but not in *C. obsidiansis*. The taxonomic distribution of these elements was uneven (Table S1) in that three species, *C. owensensis*, *C. lactoaceticus*, and *C. bescii*, contained ~80 % (65 of 81) of the copies.

Of the 68 elements that showed TIRs, 21 were assumed to be non-autonomous elements because they are either severely truncated (ORF size <70 % of IS*Cbe*4_ORF) or contain no ORF similar to IS*Cbe*4_ORF. The other 47 elements are candidate autonomous elements. The 68 elements were classified into ten families according to their DNA identity [29]. Copy numbers of families were highly uneven: two families (family A and B) include 59 of the 68 elements while the other nine were assigned into eight families. Families A and B formed two clades in the

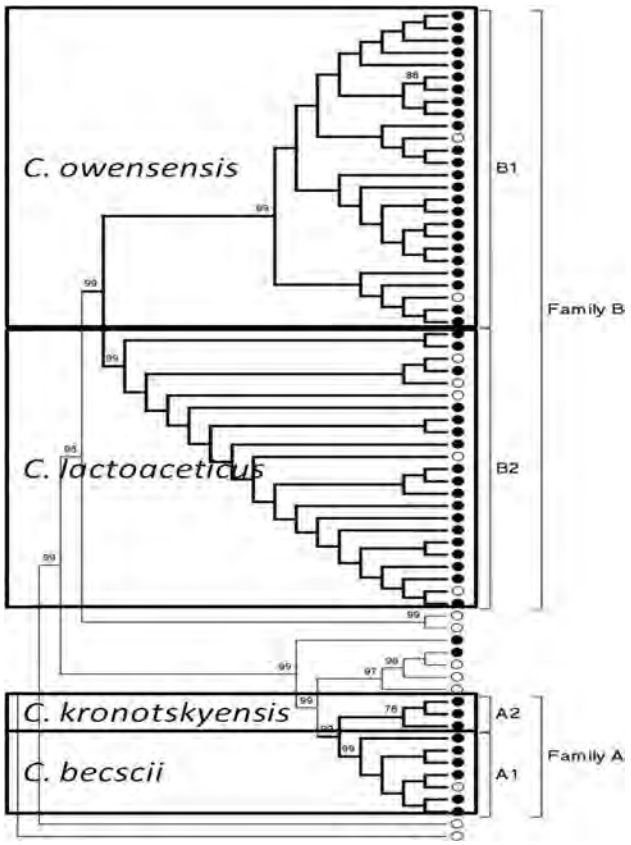

Fig. 2 Phylogeny of 68 TIR-intact elements. Only topology is presented. Bootstrap values are based on 100 replicates and are shown above branches when ≥75 %. *Black and white dots* indicate the presence or absence of TSD, respectively. When observed, all intact TSD were 8 bp in length and of >89 % identity. Host organisms of subfamilies were put in *boxes* with the species names

phylogenetic tree of the 68 elements (Fig. 2). By mapping host organisms onto the tree, families A and B could further be split into two subfamilies each: seven elements belonging to *C. bescii* formed a subclade of family A and this subclade was thus assigned as subfamily A1. Similarly, the clades containing three, 26, and 23 members from *C. kronotskyensis*, *C. owensensis* and *C. lactoaceticus* were named subfamilies A2, B1, and B2, respectively.

Variation within and between families

We built nucleotide and peptide consensus sequences of the four subfamilies and calculated identity values between them through global alignment. Nucleotide (NT) and amino acid (AA)-level identity between consensus sequences of A1 and A2 were 99 and 96 %, and between B1 and B2 were 90 and 87 %, while between family A and B were 74 and ~62 %, respectively (Table S2). Since four subfamilies were unique to specific species, they are not likely to result from lateral transfer. This facilitated the analysis of the evolutionary rate (here measured by identity: less identity indicates fast variation) of sequence change in these elements. We compared the sequence identity of subfamilies between species to that between regular orthologous genes of host genomes to see if the transposons were more variable than regular genes. A comparison of subfamilies is shown in Table S2.

As shown in Fig. 3, the distribution of identity between orthologous genes from *C. bescii* and *C. kronotskyensis* were 0.95 ± 0.08 (NT) and 0.95 ± 0.11 (AA). The degree of identity of A1 vs. A2 (NT 99 %; AA 96 %) was close to the peak values of regular genes, indicating no obvious signs of rapid TE evolution. In contrast, the mean AA identities of A1 vs. B1 (64 %) and A1 vs. B2 (60 %) elements were significantly (*p* values of 0.06 and 0.07, respectively) lower than the mean identities of orthologous genes from *C. bescii* vs. *C. owensensis* (0.88) and *C. bescii* vs. *C. lactoaceticus* (0.87). The lack of more rapid evolution in the A1 and A2 clades was unexpected, but might be a result of the very recent divergence of their host species such that the degree of variation had not yet reached a statistically significant level.

The relationships between the species *C. bescii*, *C. kronotskyensis*, *C. owensensis*, and *C. lactoaceticus* have been previously reconstructed as [((*C. bescii*, *C. kronotskyensis*), *C. owensensis*), *C. lactoaceticus*] [3]; [(*C. bescii*, *C. kronotskyensis*), (*C. lactoaceticus*, *C. owensensis*)] [2]; or [((*C. bescii*, *C. kronotskyensis*), *C. lactoaceticus*), *C. owensensis*] [30]. The first two phylogenies were based on 16s rRNA sequences while the last was from a nontraditional alignment-free method using composition vectors generated with whole gene sets. Although the phylogeny of this genus has not been well resolved, *C. bescii* was more closely related to *C. kronotskyensis* than to *C. owensensis* and *C. lactoaceticus* in all three trees. Consistent with this topology, we found that the number of strict one-to-one orthologous counterparts of *C. bescii* genes were 2,063, 120, and 69 in *C. kronotskyensis*, *C. owensensis*, and *C. lactoaceticus*, respectively. Considering these relationships between genomes, our results indicate more rapid evolution of IS*Cbe*4 and related elements than normal genes, with the higher mutation rate becoming more recognizable when examining organisms that diverged from a common ancestor at earlier dates.

Despite the many hundreds of prokaryotic genomes that have been sequenced, only 20 classes of IS elements have been found. This study uncovers the 21st class, in a clade of organisms that has received very little genomic investigation. Even when a new class of bacterial TE is discovered, the odds of finding active elements are not high, yet IS*Cbe*4 was identified as an apparently recent insertion in *C. bescii ldh*, given that *ldh* was otherwise fully intact and the presence of polymorphism for this insertion

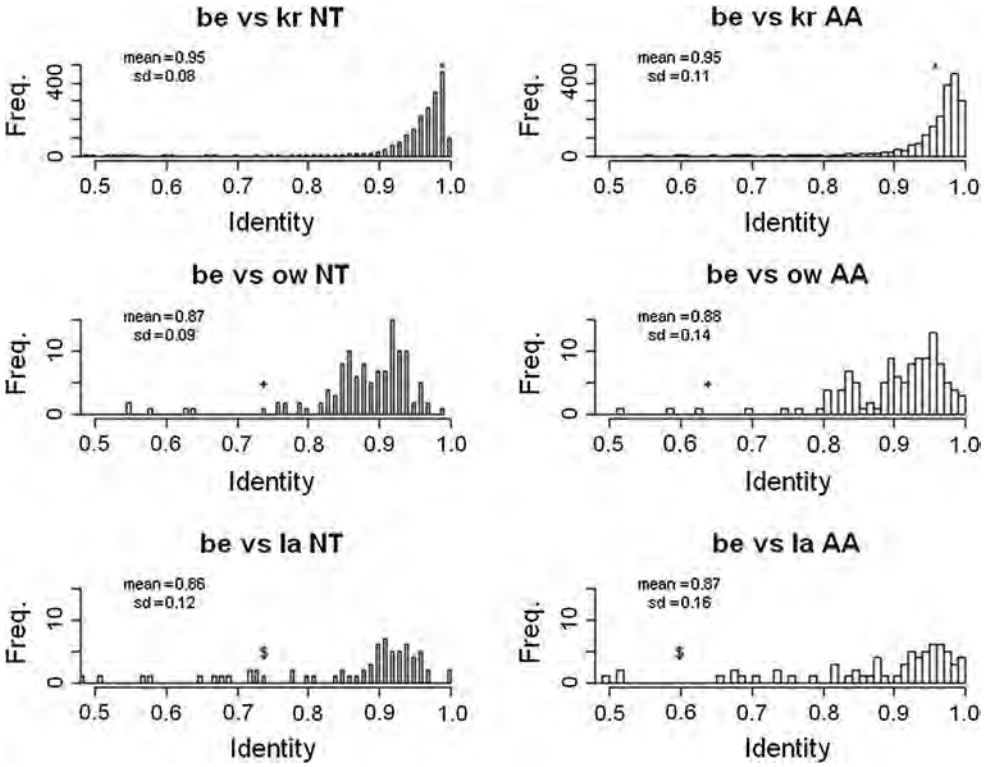

Fig. 3 Inter-species identity of regular orthologous genes and consensus sequences of transposons. Histogram of identity values between orthologous genes. *X* axis coordinates of "*", "+" and "$" mark the identity between consensus sequences of subfamilies [A1 and A2], [A1 and B1], and [A1 and B2], respectively. *Y* axis coordinates does not correspond to frequency. be, *C. bescii*; kr, *C. kronotskyensis*; ow, *C. owensensis*; la, *C. lactoaceticus*; NA, nucleotides; AA, amino acids

within *C. bescii*. In fact, the immediate parent of the strain containing the insertion was wild type at the *ldh* locus. These results indicate the value of venturing into untested waters, and we predict that additional novel TE families will be found in the thermophilic bacteria. For now, the presence and mutational activity of IS*Cbe*4 provide a new tool for *Caldicellulosiruptor* genetics and genome characterization.

Acknowledgments We thank Jennifer Copeland for outstanding technical assistance, and Robert Kelly and Sara Blumer-Schuette for providing the wild-type *Caldicellulosiruptor* species. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

References

1. Blount ZD, Grogan DW (2005) New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea. Mol Microbiol 55(1):312–325. doi:10.1111/j.1365-2958.2004.04391.x
2. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL 2nd, Adams MW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM (2012) Caldicellulosiruptor core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. J Bacteriol 194(15):4015–4028. doi:10.1128/JB.00266-12
3. Blumer-Schuette SE, Lewis DL, Kelly RM (2010) Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. Appl Environ Microbiol 76(24):8084–8092. doi:10.1128/AEM.01400-10
4. Brynestad S, Synstad B, Granum PE (1997) The *Clostridium perfringens* enterotoxin gene is on a transposable element in type A human food poisoning strains. Microbiology 143(Pt 7):2109–2115
5. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973. doi:10.1093/bioinformatics/btp348
6. Chung D, Farkas J, Westpheling J (2013) Detection of a novel active transposable element in Caldicellulosiruptor hydrothermalis and a new search for elements in this genus. J Ind Microbiol Biotechnol 40(5):517–521. doi:10.1007/s10295-013-1244-z
7. Chung D, Farkas J, Westpheling J (2013) Overcoming restriction as a barrier to DNA transformation in Caldicellulosiruptor species results in efficient marker replacement. Biotechnol Biofuels 6(1):82. doi:10.1186/1754-6834-6-82
8. Chung D, Cha M, Farkas J, Westpheling J (2013) Construction of a stable replicating shuttle vector for *Caldicellulosiruptor* species: use for extending genetic methodologies to other members of this genus. PLoS ONE 8(5):e62881
9. Collins CM, Gutman DM (1992) Insertional inactivation of an *Escherichia coli* urease gene by IS3411. J Bacteriol 174(3):883–888
10. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797. doi:10.1093/nar/gkh340

11. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J (2013) Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol 40(1):41–49. doi:10.1007/s10295-012-1202-1
12. Fetherston JD, Perry RD (1994) The pigmentation locus of *Yersinia pestis* KIM6 + is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. Mol Microbiol 13(4):697–708
13. 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(11):3760–3761. doi:10.1128/JB.00256-09
14. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120
15. Kivi M, Liu X, Raychaudhuri S, Altman RB, Small PM (2002) Determining the genomic locations of repetitive DNA sequences with a whole-genome microarray: IS6110 in *Mycobacterium tuberculosis*. J Clin Microbiol 40(6):2192–2198
16. Liyanage H, Holcroft P, Evans VJ, Keis S, Wilkinson SR, Kashket ER, Young M (2000) A new insertion sequence, ISCb1, from *Clostridium beijerinckii* NCIMB 8052. J Mol Microbiol Biotechnol 2(1):107–113
17. Maamar H, de Philip P, Belaich JP, Tardif C (2003) ISCce1 and ISCce2, two novel insertion sequences in *Clostridium cellulolyticum*. J Bacteriol 185(3):714–725
18. Mahillon J, Chandler M (1998) Insertion sequences. Microbiol Mol Biol Rev 62(3):725–774
19. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48(3):443–453
20. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314(5):1041–1052. doi:10.1006/jmbi.2000.5197
21. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16(6):276–277
22. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425
23. Siguier P, Filee J, Chandler M (2006) Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9(5):526–531. doi:10.1016/j.mib.2006.08.005
24. Stanley J, Baquar N, Threlfall EJ (1993) Genotypes and phylogenetic relationships of *Salmonella typhimurium* are defined by molecular fingerprinting of IS200 and 16S rrn loci. J Gen Microbiol 139(Pt 6):1133–1140
25. Stroeher UH, Jedani KE, Dredge BK, Morona R, Brown MH, Karageorgos LE, Albert MJ, Manning PA (1995) Genetic rearrangements in the *rfb* regions of *Vibrio cholerae* O1 and O139. Proc Natl Acad Sci USA 92(22):10374–10378
26. Svetlichnyi VA, Svetlichnaya TP, Chernykh NA, Zavarzin GA (1990) Anaerocellum-Thermophilum Gen-Nov Sp-Nov—an Extremely Thermophilic Cellulolytic Eubacterium Isolated from Hot-Springs in the Valley of Geysers. Microbiol 59(5):598–604
27. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739. doi:10.1093/molbev/msr121
28. Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev 70(2):296–316. doi:10.1128/MMBR.00048-05
29. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982. doi:10.1038/nrg2165
30. Xu Z, Hao B (2009) CV Tree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. Nucleic Acids Res 37 (Web Server issue):W174–178. doi:10.1093/nar/gkp278
31. Zverlov VV, Klupp M, Krauss J, Schwarz WH (2008) Mutations in the scaffoldin gene, *cip*A, of *Clostridium thermocellum* with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose. J Bacteriol 190(12):4321–4327. doi:10.1128/JB.00097-08