J. Mol. Microbiol. Biotechnol. (2000) 2(1): 107-113.

Fermentation Symposium
JMMB Research Article

A New Insertion Sequence, ISCb*1*, from *Clostridium beijerinckii* NCIMB 8052

Hemachandra Liyanage[2], Paul Holcroft[1], Victoria J. Evans[1], Stefanie Keis[3], Shane R Wilkinson[1,4], Eva R. Kashket[2], and Michael Young[1]*

[1]Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, SY23 3DD, UK
[2]Department of Microbiology, Boston University School of Medicine, Boston, Massachusetts 02118, USA
[3]Microbiology Department, University of Otago, PO Box 56, Dunedin, New Zealand
[4]Current Address: Department of Medical Parasitology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

Abstract

The NCIMB 8052 strain of *Clostridium beijerinckii* contains nine copies of a novel insertion sequence, ISCb*1*, belonging to the IS*4* family. The 1764 bp element has 18 bp inverted repeats at its extremities, and generates 11 bp target repeats upon insertion. It contains a 1365 bp ORF whose predicted product (455 amino acids) resembles bacterial transposases. The highly conserved DD(35)E motif is present, as are signatures characteristic of the N3 and C1 domains of bacterial transposases. Codon usage of the ORF is somewhat different from that of other *C. beijerinckii* genes, suggesting that ISCb*1* may have been acquired from another organism by horizontal gene transfer in the evolutionary past. One ISCb*1* copy lies close to the site of insertion of Tn*1545* in a mutant strain, C10, which shows a reduced tendency to degenerate (*i.e.* loss of the potential to form solvents) compared with the wild type. In the C10 strain, the characteristic pattern of DNA fragments detected by an IS-specific probe was altered, but this was due to the Tn*1545* insertion itself, rather than an ISCb*1*-mediated genome re-arrangement. There is currently no evidence that the element is involved in strain degeneration, since 12 independently isolated spontaneous mutants that had lost the ability to form solvents had the same ISCb*1* profile as that of the wild type strain. The element is apparently restricted to a series of closely related solvent-forming clostridia.

Introduction

Insertion sequences (IS) are widely distributed throughout the prokaryotic world (Mahillon and Chandler, 1998). Although phenotypically "silent" themselves, they are often associated with the acquisition of accessory functions, such as pathogenic and virulence determinants in a variety of organisms, including *Clostridium perfringens* (Brynestad *et al*., 1997; Cornillot *et al*., 1995). Insertion sequences are an important component of bacterial chromosome architecture, and are frequently involved in plasmid integration (Low, 1996) and chromosomal re-arrangements (Haack and Roth, 1995; Louarn *et al*., 1985; Savic *et al*., 1983).

Clostridium beijerinckii NCIMB 8052 is one of the solvent-forming clostridia (Jones and Woods, 1986). During laboratory batch culture with a readily used carbon source such as glucose, bacteria grow exponentially, producing mainly acetate and butyrate as fermentation end products, which results in acidification of the culture medium. Co-incident with the end of the exponential phase of growth, bacterial metabolism is re-adjusted to give neutral end products, which are mainly acetone and butanol. This, together with re-assimilation of the acetate and butyrate produced previously, permits survival for several hours, during which endospores are usually formed (Jones and Woods, 1986). During prolonged periods of laboratory culture, degenerate mutants arise that have lost the ability to form solvents and endospores (Jones and Woods, 1986; Woolley and Morris, 1990).

One potential reason for strain degeneration is genetic instability, such as loss of extra-chromosomal genetic elements, or the occurrence of deletions and/or other chromosomal re-arrangements. For example, degenerate variants of *Clostridium acetobutylicum* strains arise via loss of a megaplasmid, which encodes several of the enzymes specifically concerned with solventogenesis (Cornillot *et al*., 1997; Cornillot and Soucaille, 1996; Stim-Herndon *et al*. 1996). Gross chromosomal re-arrangements are well documented in streptomycetes, and in industrial strains they are frequently associated with degeneration of antibiotic hyper-production (Gravius *et al*., 1993). In this investigation we have characterised a novel IS from *C. beijerinckii* NCIMB 8052. The data currently available suggest that it does not play a major role in the degeneration of solvent production, to which this organism is particularly prone (Woolley and Morris, 1990).

Results

Detection of a Reiterated Sequence in *C. beijerinckii* NCIMB 8052

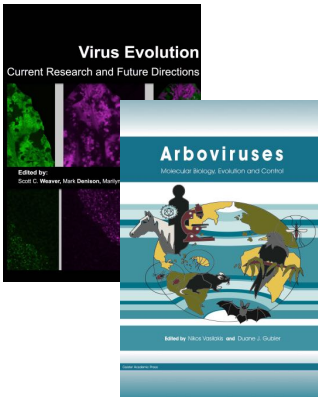
Strain C10 of *C. beijerinckii* was isolated after Tn*1545* mutagenesis of the wild type (NCIMB 8052) by screening for organisms with a reduced tendency to degenerate (Kashket and Cao, 1995). Southern hybridisation using an *aphA-3*-specific probe detected a single ca. 9.5 kbp *Hin*dIII junction fragment in strain C10 DNA (data not shown), indicating that this strain harbours a single copy of Tn*1545*. This 9.5 kbp *Hin*dIII fragment was isolated from C10 DNA and the clostridial DNA contained within it was sequenced (see experimental procedures). It encompasses the left end of Tn*1545* (including the *aphA-3* gene) and contains

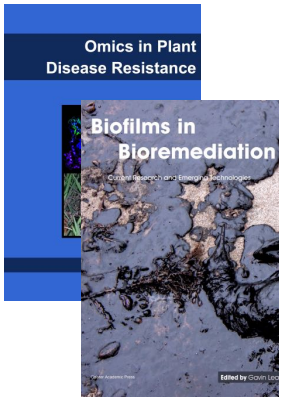
Received November 2, 1999; revised November 24, 1999; accepted November 29, 1999. *For correspondence. Email miy@aber.ac.uk; Tel. +44 (0) 1970 622348; Fax. +44 (0) 1970 622307.

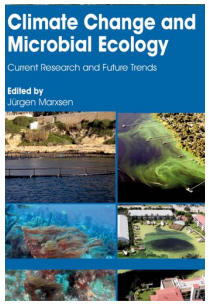
Further Reading

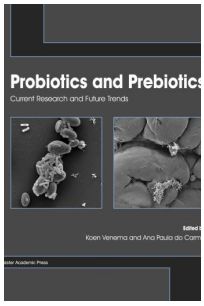
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

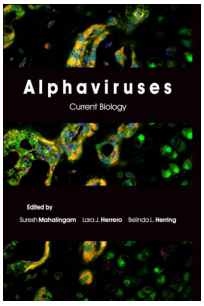
- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile

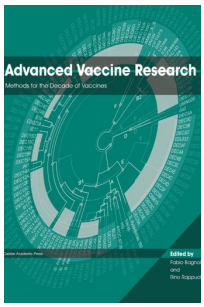


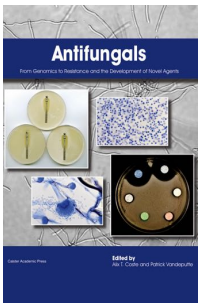












- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order

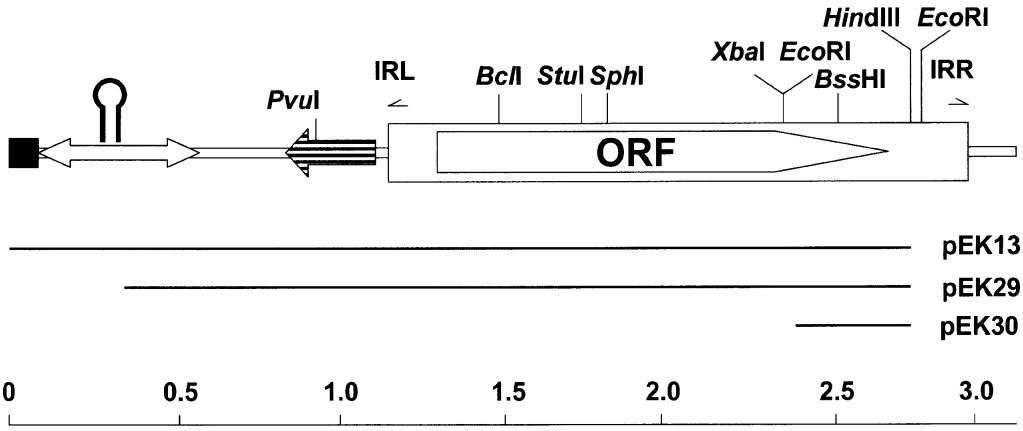


Figure 1. Diagram of the chromosomal region adjacent to the Tn*1545* insertion point in strain C10. The end of Tn*1545* (filled box), a region able to form a stable secondary structure (double headed arrow), and the *sinR*-like gene (arrow) are shown, as is the IS (open box) together with its associated ORF and terminal inverted repeats (IRL and IRR). The positions of landmark restriction sites are indicated, as are the inserts cloned in plasmids pEK13, pEK29 and pEK30. The scale below the diagram is in kbp.

2638 bp of flanking chromosomal DNA. Immediately adjacent to the transposon end is a 486 nt segment that could potentially form a very stable secondary structure (ΔG = 431 kJ/mole-data not shown) followed by two ORFs (Figure 1). The predicted (92 aa) product of the first ORF resembles *Bacillus subtilis* SinR (37% identity and 57% similarity over a 51 residue overlap) and the product of an ORF from *Staphylococcus aureus* NCTC 8325 (48% identity and 72% similarity over an 88 residue overlap). The predicted product of the second ORF (455 aa) shows slight similarity to bacterial transposases, including those of IS*1384* from *Pseudomonas putida* (15% identity and 26% similarity with g2995640 over a 350 residue overlap) and IS*1194* from *Streptococcus thermophilus* (17% identity and 24% similarity with g2665345 over a 381 residue overlap).

When employed as a hybridisation probe, the 9.5 kbp *Hin*dIII fragment detected multiple *Hin*dIII fragments from the chromosomes of both the wild type and the C10 strain (Figure 2A). The uppermost of the seven bands detected had a reduced mobility in C10 DNA. This altered band corresponds to the cloned *Hin*dIII fragment that was targeted by Tn*1545* insertion and detected by the *aphA-3* probe (see above). An identical pattern of hybridisation was obtained when a 378 bp *Eco*RI-*Hin*dIII sub-fragment from the Tn*1545*-distal end of the cloned DNA (Figure 1) was employed as probe (Figure 2B). These results indicated that a reiterated element lay close to the site of Tn*1545* insertion in the C10 strain.

Determination of the Size of the Reiterated Sequence

It seemed most likely that the reiterated element would encompass at least the larger ORF shown in Figure 1 and evidence supporting this contention was obtained in ligation-anchored PCR experiments. DNA from the wild type strain was digested with *Hin*dIII and then ligated with an equivalent quantity of *Hin*dIII-digested pMTL20 (Chambers *et al.*, 1988). The ligation mixture was employed as template using as primers either 29S or 29S2 together with either the M13 universal or the M13 reverse primer. After PCR amplification, a family of between four and six amplicons was obtained, derived from the various *Hin*dIII fragments seen in Figure 2. All the amplicons were reduced in size by about 150, 230 and 490 bp after digestion with either *Sph*I, or *Stu*I or *Bcl*I, respectively, but not after digestion with *Pvu*I (data not shown). These results established that the reiterated sequence extends through the *Sph*I, *Stu*I and *Bcl*I sites that are within the large ORF, but not as far as the *Pvu*I site within the *sinR*-like gene (see Figure 1).

To pinpoint the ends of the reiterated element more precisely, DNA samples from the wild type strain were digested with either *Hin*dIII (left end) or *Xba*I (right end) and ligated with similarly digested pBluescriptKS$^+$ (to provide the second priming site). Using a combination of the KS primer together with either DWL (left end) or DW2

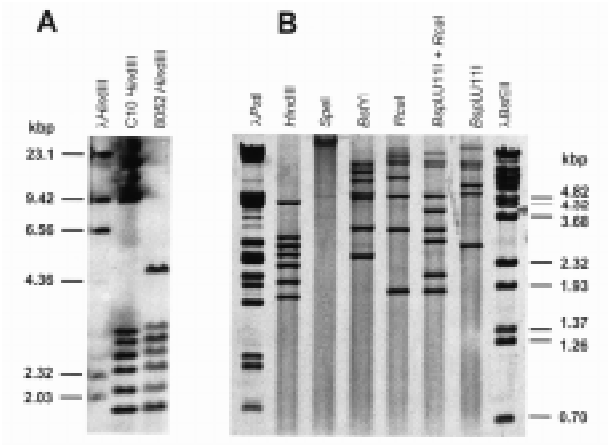


Figure 2. A reiterated sequence in the *C. beijerinckii* chromosome. In part A, genomic DNA from strains NCIMB 8052 and C10 was digested with *Hin*dIII and probed with a sub-fragment of the insert from pEK29 (see Figure 1), lacking the terminal 378 bp *Eco*RI-*Hin*dIII segment. In part B, *C. beijerinckii* NCIMB 8052 DNA was cleaved with a range of restriction enzymes and probed with pEK30 DNA. The size markers are bacteriophage lambda DNA digested with *Hin*dIII (Part A, lane 1) *Pst*I and *Bst*EII (Part B lanes 1 and 8). (The 4.36 kbp *Hin*dIII fragment in Part A, lane 1, is under-represented owing to its interaction with the 23.1 kbp fragment during electrophoresis.)

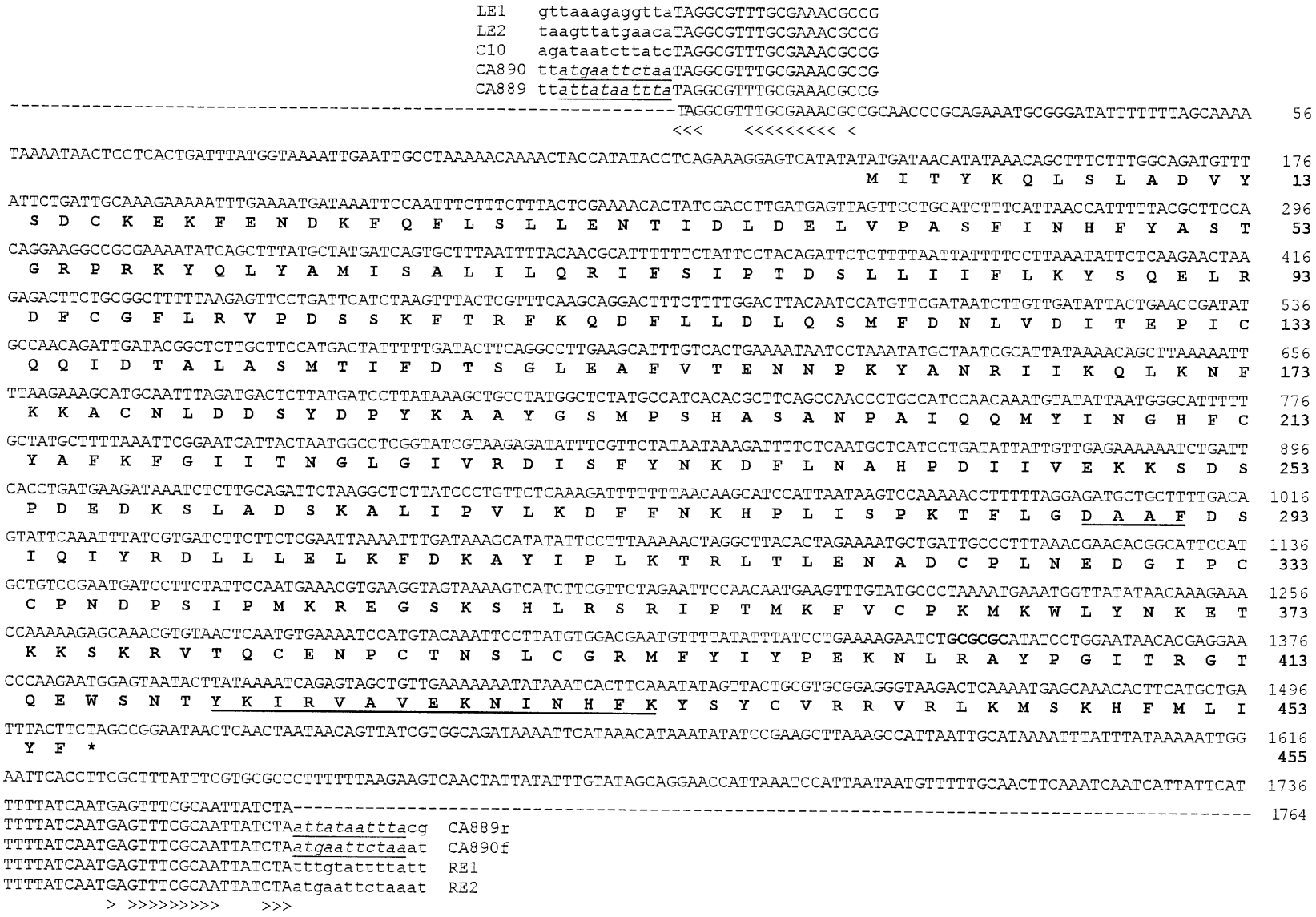


Figure 3. Nucleotide sequence of ISCb*1*. DNA fragments encompassing the ends of ISCb*1* were obtained using either ligation-anchored PCR (LE1, LE2, RE1, RE2) or inverse PCR (CA889, CA890) with wild type DNA as template. The sequences of these segments were identical to that of the corresponding segment of ISCb*1*, as shown in the Figure (nt 1-103 for LE1 and LE2; nt 1-524 and 1673-1764 for CA889; nt 1-518 and 1673-1764 for CA890; nt 1734-1764 for RE1 and RE2). They are aligned in the Figure together with the left end of the element from strain C10 DNA (nt 1-1577). The nucleotide sequence of ISCb*1* is in upper case with the rare *Bss*HII site (nt 1348-1353) in bold, and the flanking sequences in lower case. The 18 bp inverted repeats at the ends of the element are indicated below ISCb*1* and the 11 nt duplications of the target sequences in DNA from clones CA889 and CA890 are underlined. The predicted product of the ORF is shown below the DNA sequence in upper case bold, using the single letter amino acid code, and the segments encompassing the conserved N3 and C1 motifs of bacterial transposases (see text) are underlined.

(right end), multiple amplicons were obtained. Two amplicons from each reaction were cloned and sequenced. The sequences of the two DWL amplicons (called LE1 and LE2) were compared with that of the insert in pEK13 (Tn*1545* junction fragment from strain C10). All three sequences were identical from the DWL priming site to co-ordinate 1145 in Figure 1. Thereafter, their sequences diverged. The sequences of the two DW2 amplicons (called RE1 and RE2) diverged at a point 187 nt beyond the *Hin*dIII site at the extremity of the insert in pEK13 (see Figure 1). These data, effectively fixing the locations of both extremities of the reiterated element, are summarised in Figure 3.

The Reiterated Sequence is an Insertion Sequence

The sequence data obtained above revealed that the reiterated element has a size of 1764 bp and that it has imperfect 18 bp inverted repeats (13/18 matches) at its extremities (Figure 3). The 1365 bp ORF with weak similarity to bacterial transposases (see above) lies between bases 140 and 1504. These features suggested that the reiterated element is an insertion sequence (IS). We therefore determined whether there are duplications of host sequences on either side of the element in the bacterial chromosome.

Inverse PCR was employed to isolate the host DNA sequences immediately adjacent to two different copies of the element in the bacterial genome. The putative IS lacks sites for *Rca*I and *Bsp*LU11I, both of which have A + T-rich target sequences and cleave *C. beijerinckii* DNA frequently. Target DNA from the wild type strain was digested with these enzymes and their compatible ends ligated at a low DNA concentration, to favour the generation of circular products. The ligation mixture was employed as template for inverse PCR using primers DWAL1 and DWAR1, which prime DNA synthesis outwards towards the ends of the element. Multiple amplicons were obtained two of which were cloned and sequenced. These data (see CA889 and CA890 in Figure 3) revealed the presence of an 11 bp duplication on either side of the element confirming that it is an IS. They also confirmed the previously deduced extremities of the element. This is the first example of an IS from *C. beijerinckii*, and it has been denoted ISCb*1*.

Analysis of the DNA sequences flanking ISCb*1* in CA890 (accession number AJ250470) revealed that it has disrupted an ORF that could encode a response regulator of the AraC/XylS family (33% identity and 57% similarity to *B. subtilis yfiF* over an 82 residue overlap). The DNA sequences flanking ISCb*1* in CA889 (accession number AJ250471) probably correspond to an intergenic, non-coding region of the bacterial chromosome

C. beijerinckii* NCIMB 8052 Contains at Least Nine Copies of ISCb*1

Inspection of the data shown in Figure 2 suggests that two of the seven bands corresponding to *Hin*dIII fragments bearing ISCb*1* may be doublets. Their hybridisation intensity is enhanced compared with that of the other bands. Southern hybridisation experiments were carried out using DNA digested with a variety of enzymes that cleave clostridial DNA frequently but do not cleave ISCb*1*. Digestion with *Bsp*LU11I + *Rca*I gave nine hybridising fragments (Figure 2B) indicating that there are at least nine copies of ISCb*1* in *C. beijerinckii* NCIMB 8052. Their

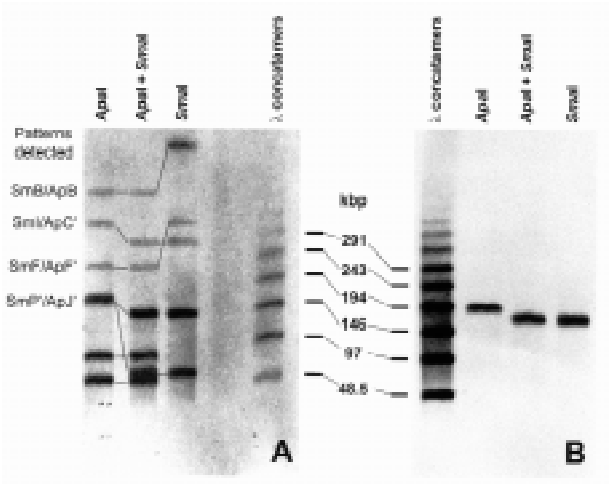


Figure 4. Detection of ISCb*1* in *Apa*I and *Sma*I macro-restriction fragments from *C. beijerinckii* DNA. Macro-restriction fragments were prepared from strains NCIMB 8052 (Part A) and C10 (Part B) as described previously (Wilkinson and Young, 1995). An ISCb*1*-specific probe (pEK30) was employed in part A and an *aphA-3*-specific probe (see Experimental procedures) in part B. Bands corresponding to macro-restriction fragment profiles that had been identified previously (Wilkinson and Young, 1995) are indicated. Note that the SmP'/ApJ' profile detected in part B has been shifted relative to its counterpart in part A, owing to the presence of a copy of Tn*1545* (25.3 kbp - Caillaud *et al*., 1987a) in the hybridising fragments.

locations in the bacterial chromosome were investigated by probing *Sma*I and *Apa*I macro-restriction fragments of the wild type strain with an ISCb*1*-specific probe. This experiment (Figure 4A) revealed a complex pattern of hybridising fragments corresponding to four of the profiles previously reported, together with at least three additional profiles not previously described (Wilkinson and Young,

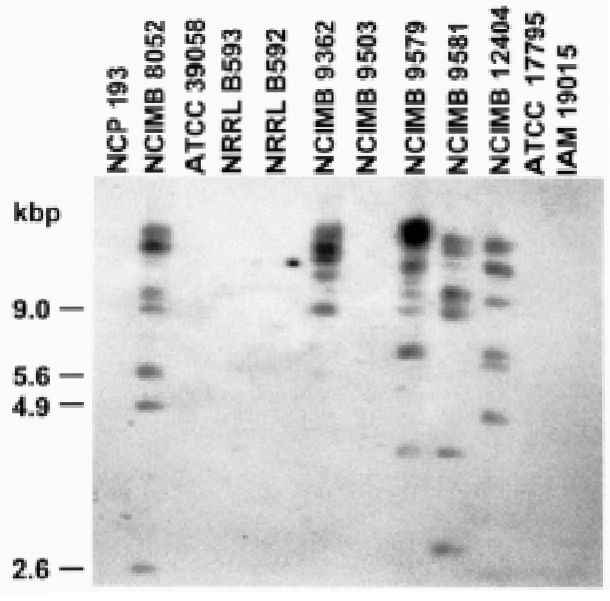


Figure 5. Distribution of ISCb*1* among *C. beijerinckii* strains representing twelve different DNA fingerprint groups. DNA samples were digested with *Bsp*LU11I, and probed with pEK30 DNA.

Table 1. Comparison of the G + C Contents of the Three Codon Positions in the ISCb*1* ORF with those for Other Genes from *Clostridium beijerinckii* NCIMB 8052 and *Clostridium acetobutylicum* ATCC 824

	C. beijerinckii	*C. acetobutylicum*	IS ORF
Coding GC%	32.9	33.1	33.9
1st letter GC%	44.9	45.5	41.9
2nd letter GC%	35.6	34.5	34.0
3rd letter GC%	18.2	19.3	25.9

1995). One of the profiles detected is SmP'/ApJ' (Figure 4A). A modified form of this profile was detected when strain C10 DNA was hybridised with an *aphA-3*-specific probe (Figure 4B). In strain C10, the macro-restriction fragments detected are about 25 kbp larger, since they contain Tn*1545* (Caillaud *et al*., 1987a).

The distribution of ISCb*1* amongst other *C. beijerinckii* strains was determined. One representative strain was chosen from each of the five different fingerprint groups of Keis *et al*. (1995), together with seven other *C. beijerinckii* strains which represent a further seven DNA fingerprint groups (Keis, 1996). The element was detected in 5 of the 12 strains tested (Figure 5). A more extensive survey was made of the eight strains comprising fingerprint group 6 (*viz* NCIMB 6445, NCIMB 8049, NCIMB 8052, NCIMB 8653, NCIMB 10132, NRRL B591, NRRL B594 and 214 (Keis *et al*., 1995)). The patterns of hybridisation after digestion of the DNA with either *Hin*dIII or *Bsp*LU11I were all identical (data not shown). The remaining four strains in which the element was present are either the sole representatives of their fingerprint group (NCIMB 9581 and NCIMB 12404) or one of two strains comprising their group (NCIMB 9362, NCIMB 9579). Multiple copies of ISCb*1* were present in each of the five strains representing five different fingerprint groups in Figure 5. Although each representative strain showed a distinct hybridisation profile (using *Bsp*LU11I-digested DNA), there appeared to be common bands in some profiles (Figure 5).

Table 2. Primers Employed for PCR and Sequencing

Primer	Sequence (5' → 3')[a]	Co-ordinates[b] or Reference
29R	GGAAGGAGGATTGTAAAATG	696 → 715
29R2	GAGCACCAAGCAATA	1015 → 1029
29S	GTCAAAAGCAGCATCTCC	2159 → 2142
29S2	CCGAGGCCATTAGTA	1960 → 1946
C10P1	CCTAACTCGGTAGACACTTGCAT	234 → 212
C10P2	ACTGTCTACTGTCTACTAAGC	288 → 268
DWAL1	CTGTTGGCATATCGGTTCAG	1688 → 1669
DWAR1	TTAAGAA*GTCgAC*TATTATATTTG	2794 → 2817
DW2	ATCCATGTACAAATTCCTTATGT	2434 → 2456
DWL	at*cctgcag*AAATTGCATGCTTTCTTAAA	1819 → 1800
DWL3	AGCCGCAGAAGTCTCTTAGTT	1575 → 1555
DWL4	TCATATATATGACTCCTTTCTGAG	1287 → 1264
TNLE	GGATAAATCGTCGTATCAAAG	(Trieu-Cuot *et al*., 1991)

[a]Lower case letters indicate nt changes introduced to create restriction sites (italics)
[b]numbers refer to positions in Figure 1

Does ISCb*1* Provoke Degeneration in *C. beijerinckii* NCIMB 8052?

A small collection of degenerate mutants was obtained by plating the wild type strain to give a low colony density on CBM (O'Brien and Morris, 1971) and purifying organisms from colony outgrowths, as described previously (Kashket and Cao, 1995). DNA was isolated from twelve independent mutants, digested with *Rca*I + *Bsp*LU11I, separated electrophoretically, blotted and probed with an ISCb*1*-specific probe. The hybridisation patterns of all strains were identical to that shown in Figure 2B, indicating that ISCb*1*-mediated genome rearrangements do not make a major contribution to strain degeneration on agar-solidified CBM.

Discussion

The results reported in this communication indicate that *C. beijerinckii* NCIMB 8052 and closely related strains contain a novel 1764 bp IS, which we have called ISCb*1*, that appears to belong to the IS*4* family (Mahillon and Chandler, 1998; Rezsöhazy *et al*., 1993). The predicted 455 aa product of the single ORF encoded by the element has weak similarity to bacterial transposases from IS*1194* (*S. thermophilus*) and IS*1384* (*P. putida*). Moreover, residues 288-291 (**D**A**AF**) and 421-435 (**Y**KI**R**VA**VE** KNINHF**K**) (Figure 3) correspond to the **D**(1)[**G**/**A**][**Y**/**F**] and [**Y**/**L**](2)**R**(2)[**I**/**L**/**V**]**E**(6)**K** signatures characteristic of the N3 and C1 domains of bacterial transposases (Rezsöhazy *et al*., 1993). The underlined residues in these signatures are part of a highly conserved **DD**(35)**E** catalytic triad (Mahillon and Chandler, 1998). One of several aspartic acid residues located upstream from the N3 signature (Figure 3) could potentially correspond to the first residue of the catalytic triad. The *C. beijerinckii* element apparently generates 11 bp repeats of its target sequence, which is within the range of 9-12 bp reported for the IS*4* family (Mahillon and Chandler, 1998). The ends of the element have imperfect, 18 bp inverted repeats (13/18 bases matching), with a terminal A residue, which is unusual for IS*4* family members. Curiously, the base compositions of the left and right ends are very dissimilar, with 24/40 and 9/40 bases being G or C at the left and right ends, respectively. The G+C-rich sequences at the left end form a family of short G+C-rich inverted repeats, which are commonly encountered at the ends of insertion sequences and transposons (Mahillon and Chandler, 1998).

The IS has an overall G+C content of 33.2 mole percent. In the solventogenic clostridia, intergenic regions tend to be even more A+T-rich than coding regions (Young *et al*., 1989), whereas both coding and non-coding regions of ISCb*1* have very similar G+C contents. Although the overall G+C content of the coding region does conform to that of other *C. beijerinckii* (and *C. acetobutylicum*) genes, there are some interesting differences in codon usage. The first codon position has a higher proportion of A+U residues and this is compensated by a lower proportion of A+U residues in the third codon position. This is due, in large measure, to an increased abundance of several codons ending in C, *viz*: CUC_{leu} = 14% (6/23); GGC_{gly} = 36% (5/14); UUC_{Phe} = 36% (11/31); UGC_{cys} = 69% (9/13). (These codons are used at the following frequencies in *C. beijerinckii* and *C. acetobutylicum*, respectively: CUC_{leu} = 2% and 3%; GGC_{gly} = 8% and 9%; UUC_{Phe} = 19% and

18%; UGC_{cys} = 24% and 29%.) These differences in codon usage, together with the higher than average G + C content of the non-coding regions, suggest that ISCb*1* may have been acquired from another organism by horizontal gene transfer in the evolutionary past.

The solventogenic clostridia are a heterogeneous assemblage of organisms (Keis *et al.*, 1995; Wilkinson and Young, 1993; Wilkinson *et al.*, 1995) and ISCb*1* shows a rather restricted distribution among them (Figure 5). A search of the completed sequence of the *C. acetobutylicum* ATCC 824 genome (http://www.genomecorp.com/sequence_center/bacterial_genomes/) revealed that ISCb*1* is not present in this organism. It is apparently confined to a subset of *C. beijerinckii* – like strains, i.e. biotype/DNA fingerprint group 6 of Keis *et al.* (1995), together with four other fingerprint groups, each containing one or two strains (Keis, 1996). It is significant in this context that the element contains a rare *Bss*HII restriction site (Figure 1). This was one of the enzymes employed for genomic fingerprinting by Keis *et al.* (1995) and ISCb*1* therefore contributed to their scheme for discriminating strains. All eight members of group 6 had identical genetic fingerprints (Keis *et al.*, 1995) and they all showed identical ISCb*1* profiles. The four other strains in which the element was found (viz. NCIMB 9362, NCIMB 9579, NCIMB 9581 and NCIMB 12404) showed different ISCb*1* profiles (Figure 5). Although the number of discrete bands varied from one strain to another, some common bands appeared to be present. The fact that these strains show different ISCb*1* profiles constitutes circumstantial evidence of the occurrence IS-mediated genome re-arrangements in the comparatively recent evolutionary past. Experiments are currently in progress to determine whether the element is capable of transposition in *C. beijerinckii* NCIMB 8052.

Four of the nine copies of ISCb*1* have been located on the *C. beijerinckii* NCIMB 8052 chromosome in hybridisation experiments with *Sma*I and *Apa*I macro-restriction fragments. It should now be possible further to refine the physical map of the bacterial chromosome (Wilkinson and Young, 1995) using the enzyme *Bss*HII, together with IS-specific probes, to determine the precise positions and orientations of all nine copies of ISCb*1*.

The reason why strain C10 shows a reduced tendency to degenerate is not yet understood. Analysis of the sequence adjacent to the right end of the transposon failed to reveal the presence of an open reading frame that would be disrupted by Tn*1545* insertion. However, were this 486 bp region transcribed, the resulting RNA would form a very stable hairpin structure (ΔG = -431 kJ/mole). It is tempting to speculate that this might correspond to a regulatory RNA molecule, whose disruption in the mutant is associated with a reduced tendency to degenerate. Further exploration of this hypothesis will necessitate isolation of the entire gene from the wild type strain and characterisation of the phenotype that results when it is replaced by a mutated derivative generated *in vitro*.

Experimental Procedure

Bacterial Strains and Growth Media

Clostridial strains were grown using either medium T.5, which is a modified (Kashket and Cao, 1993) form of medium TYA (Ogata and Hongo, 1973) containing 0.5% (w/v) glucose, or CBM (O'Brien and Morris, 1971), or supplemented tryptone-yeast extract-glucose medium as described previously (Keis *et al.*, 1995), solidified with 1.2 % agar, as appropriate. Erythromycin (25 μg/ml) was incorporated into the growth medium for strain C10 of *C. beijerinckii* (Kashket and Cao, 1995), which harbours a single copy of Tn*1545*. LB medium supplemented with ampicillin (100 μg/ml) was employed to select and grow *E. coli* DH5α strains harbouring recombinant pMTL20 and pBluescript KS$^+$ plasmids.

DNA Manipulations

Standard procedures (Sambrook *et al.*, 1989) were employed for all DNA manipulations. The Tn*1545* left end junction fragment was isolated from strain C10 by electrophoretic separation of *Hin*dIII digested DNA, purification of fragments in the 8.5 – 10 kbp range and cloning in pBluescript KS$^+$. Plasmid pEK13 containing the desired fragment, which includes the Tn*1545 aphA-3* gene, was obtained by selecting kanamycin-resistant transformants of *E. coli* DH5α. The clostridial DNA within pEK13 was sequenced using the dideoxy chain termination procedure (Sanger *et al.*, 1977) with various plasmid sub-clones, some of which are shown in Figure 1, in combination with several of the primers listed in Table 2. Sequence assembly and analysis was undertaken using the GCG suite of programs (Devereux *et al.*, 1984) (Wisconsin Package Version 9.1, Genetics Computer Group, Madison, Wisconsin) at the UK Daresbury laboratory. Database comparisons were undertaken at the National Centre for Biotechnology Information web site (http://www.ncbi.nlm.nih.gov/) using BLAST, gapped BLAST and psi-BLAST (Altschul *et al.*, 1990; Altschul *et al.*, 1997). Analysis of codon usage was undertaken at the Kazusa DNA Research Institute web site (http://www.kazusa.or.jp/codon/countcodon.html) using data corresponding to 126 coding sequences from *C. acetobutylicum* (41861 codons), 14 coding sequences from *C. beijerinckii* (4537 codons) and the 455 codons corresponding to the ISCb*1* ORF. The entire *C. acetobutylicum* genome sequence has now been determined (http://www.genomecorp.com/sequence_center/bacterial_genomes/), but annotation is still in progress and the full complement of gene products is not therefore currently available for codon usage analysis.

The various primers employed for sequencing and for PCR are listed in Table 2. PCR products were cloned in either pMTL20 (Chambers, *et al.*, 1988) or pBluescript KS$^+$ (Stratagene).

Ligation-Anchored PCR

Ligation-anchored PCR was employed to isolate fragments LE1 and LE2, containing the left end of two different copies of ISCb*1*. *Hin*dIII fragments of wild type DNA were ligated with *Hin*dIII-digested pBluescript KS$^+$ and employed for PCR using primers DWL and KS. Between five and seven amplicons of different sizes were obtained in different experiments. Two amplicons were isolated, digested with *Hin*dIII + *Pst*I, cloned in *Hin*dIII + *Pst*I-digested pBluescript KS$^+$ and sequenced, using primers DWL, DWL3 and DWL4. A similar method was employed to isolate fragments RE1 (950 bp) and RE2 (850 bp), containing the right end of ISCb*1*. *Xba*I fragments of wild type DNA were ligated with *Xba*I-digested pBluescript KS$^+$ and the ligation mixture was employed for PCR using primers, DW2 and KS. Five or more amplicons of different sizes were obtained in different experiments. Two of them were isolated, digested with *Hin*dIII + *Xba*I, cloned into *Hin*dIII + *Xba*I-digested pBluescript KS$^+$ and sequenced using the standard M13 reverse and KS primers.

Inverse PCR

The ends of two different copies of ISCb*1* together with flanking DNA were isolated using inverse PCR. Wild type DNA was digested with *Rca*I + *Bsp*LU11I, self-ligated at a low DNA concentration and subjected to PCR using primers DWAL1 and DWAR1. About five amplicons were generated. Two of them were cloned in pMTL20 and sequenced using the standard M13 universal and reverse primers.

Southern Hybridisations

DNA fragments were transferred to Hybond N membranes (Amersham) and hybridised with digoxigenin-labelled probes (Boehringer) using standard procedures. Plasmid pAT187 (Trieu-Cuot *et al.*, 1987), which contains an *aphA-3* gene very similar to that of Tn*1545* (Caillaud *et al.*, 1987b), was employed as an *aphA-3*-specific probe. Plasmid pEK30 (Figure 1) was employed as an ISCb*1*-specific probe. In all hybridisations, a small quantity of digoxigenin-labelled phage λ DNA was included to detect the size markers. Macro-restriction fragments of clostridial DNA were prepared as described previously (Wilkinson and Young, 1995).

Nucleotide Sequence Accession Numbers

The nucleotide sequences reported in this paper have been deposited in the EMBL database under the following accession numbers: AJ250468 (ISCb*1*), AJ250469 (sequences between the left end of Tn*1545* and the left end of ISCb*1* in the C10 strain), AJ250470 and AJ250471 (sequences flanking two copies of ISCb*1* from the wild type strain).

Acknowledgements

We are grateful to Patrick Tidball and Graham Price for technical assistance and to David Jones for helpful discussions concerning strain comparisons. Work in the laboratories of MY and ERK was supported by grant aid from the UK Biotechnology and Biological Sciences Research Council (T04480) and the Co-operative State Research, Education, and Extension Service, U.S. Department of Agriculture, under Agreement No. 97-35504-5143.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

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

Brynestad, S., Synstad, B. and Granum, P.E. 1997. The *Clostridium perfringens* enterotoxin gene is on a transposable element in type A human food poisoning strains. Microbiology 143: 2109-2115.

Caillaud, F., Carlier, C., and Courvalin, P. 1987a. Physical analysis of the conjugative shuttle transposon Tn*1545*. Plasmid 17: 58-60.

Caillaud, F., Trieu-Cuot, P., Carlier, C. and Courvalin, P. 1987b. Nucleotide sequence of the kanamycin resistance determinant of the pneumococcal transposon Tn*1545*: evolutionary relationships and transcriptional analysis of *aphA-3* genes. Mol. Gen. Genet. 207: 509-513.

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. 1988. The pMTL *nic-* cloning vectors. 1. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. Gene 68: 139-149.

Cornillot, E., Nair, R.V., Papoutsakis, E.T. and Soucaille, P. 1997. The genes for butanol and acetone formation in *Clostridium acetobutlicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. J. Bacteriol. 179: 5442-5447.

Cornillot, E., Saint-Joanis, B., Daube, G., Katayama, S., Granum, P.E., Canard, B. and Cole, S.T. 1995. The enterotoxin gene (*cpe*) of *Clostridium perfringens* can be chromosomal or plasmid-borne. Mol. Microbiol. 15: 639-647.

Cornillot, E. and Soucaille, P. 1996. Solvent-forming genes in clostridia. Nature 380: 489.

Devereux, J., Haeberli, P. and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12: 387-395.

Gravius, B., Bezmalinovic, T., Hranueli, D. and Cullum, J. 1993. Genetic instability and strain degeneration in *Streptomyces rimosus*. Appl. Environ. Microbiol. 59: 2220-2228.

Haack, K.R. and Roth, J.R. 1995. Recombination between chromosomal IS*200* elements supports frequent duplication formation in *Salmonella typhimurium*. Genetics 141: 1245-1252.

Jones, D.T. and Woods, D.R. 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. 50: 484-524.

Kashket, E.R. and Cao, Z.Y. 1993. Isolation of a degeneration-resistant mutant of *Clostridium acetobutylicum* NCIMB 8052. Appl. Environ. Microbiol. 59: 4198-4202.

Kashket, E.R. and Cao, Z.Y. 1995. Clostridial strain degeneration. FEMS Microbiol. Rev. 17: 307-315.

Keis, S. 1996. Taxonomy and phylogeny of industrial solvent-producing clostridia. Ph.D thesis, University of Otago, Dunedin, New Zealand.

Keis, S., Bennett, C.F., Ward, V.K. and Jones, D.T. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. Int. J. Syst. Bacteriol. 45: 693-705.

Louarn, J.M., Bouche, J.P., Legendre, F., Louarn, J. and Patte, J. 1985. Characterization and properties of very large inversions of the *E. coli* chromosome along the origin-to-terminus axis. Mol. Gen. Genet. 201: 467-476.

Low, K.B. 1996. Hfr strains of *Escherichia coli* K-12. In *Escherichia coli* and *Salmonella:* cellular and molecular biology, 2nd edition. Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C. , Low, K.B. , Magasanik, B., Reznikoff, W.S. , Riley, M., Schaecter, M. and Umbarger, H.E. eds. ASM Press, Washington, D. C. p. 2402-2412.

Mahillon, J. and Chandler, M. 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62: 725-774.

O'Brien, R.W. and Morris, J.G. 1971. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. J. Gen. Microbiol. 68: 307-318.

Ogata, S. and Hongo, M. 1973. Bacterial lysis of *Clostridium* species. I. Lysis of *Clostrdium* species by univalent cations. J. Gen. Appl. Microbiol. 19: 251-261.

Rezsöhazy, R., Hallet, B., Delcour, J. and Mahillon, J. 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. 9: 1283-1295.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning, A Laboratory Manual, Second edn. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74: 5463-5467.

Savic, D.J., Romac, S.P. and Ehrlich, S.D. 1983. Inversion in the lactose region of *Escherichia coli* K-12: inversion termini map within IS*3* elements alpha 3 beta 3 and beta 5 alpha 5. J. Bacteriol. 155: 943-946.

Stim-Herndon, K.P., Nair R., Papoutsakis E.T., and Bennett G.N. 1996. Analysis of degenerate variants of *Clostridium acetobutylicum* ATCC 824. Anaerobe 2: 11-18.

Trieu-Cuot, P., Carlier, C., Martin, P. and Courvalin, P. 1987. Plasmid transfer by conjugation from *Escherichia coli to* Gram-positive bacteria. FEMS Microbiol. Lett. 48: 289-294.

Trieu-Cuot, P., Carlier, C., Poyart-Salmeron, C. and Courvalin, P. 1991. An integrative vector exploiting the transposition properties of Tn*1545* for insertional mutagenesis and cloning of genes from Gram-positive bacteria. Gene 106: 21-27.

Wilkinson, S.R. and Young, M. 1993. Wide diversity of genome size among different strains of *Clostridium acetobutylicum*. J. Gen. Microbiol. 139: 1069-1076.

Wilkinson, S.R. and Young, M. 1995. A physical map of the *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052 chromosome. J. Bacteriol. 177: 439-448.

Wilkinson, S.R., Young, M., Goodacre, R., Morris, J.G., Farrow, J.A.E. and Collins, M.D. 1995. Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum*. FEMS Microbiol. Lett. 125: 199-204.

Woolley, R.C. and Morris, J.G. 1990. Stability of solvent production by *Clostridium acetobutylicum* in continuous culture: strain differences. J. Appl. Bacteriol. 69: 718-728.

Young, M., Staudenbauer, W.L. and Minton, N.P. 1989. Genetics of *Clostridium*. In Clostridia. Biotechnology Handbooks. N. P. Minton and D. J. Clarke, eds. Plenum, New York. p. 63-103.