(12) STANDARD PATENT (11) Application No. **AU 2008251465 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Gene knockout mesophilic and thermophilic organisms, and methods of use thereof

(51) International Patent Classification(s)
C07H 21/02 (2006.01) ***C12N 15/00*** (2006.01)
C12N 1/20 (2006.01)

(21) Application No: **2008251465** (22) Date of Filing: **2008.05.09**

(87) WIPO No: **WO08/141174**

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/916,978	**2007.05.09**	**US**

(43) Publication Date: **2008.11.20**
(44) Accepted Journal Date: **2013.01.10**

(71) Applicant(s)
Mascoma Corporation

(72) Inventor(s)
Hogsett, David A.;Rajgarhia, Vineet B.

(74) Agent / Attorney
Shelston IP, L 21 60 Margaret St, Sydney, NSW, 2000

(56) Related Art
LYND L R ET AL: "MICROBIAL CELLULOSE UTILIZATION: FUNDAMENTALS AND BIOTECHNOLOGY" MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 66, no. 3,
1 September 2002, pages 506-577, ISSN: 1092-2172
DESAI S G ET AL: APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, SPRINGER VERLAG,
BERLIN, DE,vol. 65, no. 5, 6 March 2004, pages 600-605, ISSN: 0175-7598
WO 2007/027828 A2 (PHYLLOM LLC) 8 March 2007

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 November 2008 (20.11.2008)

PCT

(10) International Publication Number
WO 2008/141174 A3

(51) International Patent Classification:
C07H 21/02 (2006.01) *C12N 15/00* (2006.01)
C12N 1/20 (2006.01)

(21) International Application Number:
PCT/US2008/063237

(22) International Filing Date: 9 May 2008 (09.05.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/916,978 9 May 2007 (09.05.2007) US

(71) Applicant *(for all designated States except US)*: **MASCOMA CORPORATION** [US/US]; 1380 Soldiers Field Road, Boston, MA 02135 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HOGSETT, David, A.** [US/US]; Six Winter Hill, Grantham, NH 03753 (US). **RAJGARHIA, Vineet, B.** [US/US]; 21 Spencer Street #201, Lebanon, NH 03766 (US).

(74) Agents: **GORDON, Dana, M.** et al.; Foley Hoag LLP, Seaport World Trade Center West, 155 Seaport Blvd., Boston, MA 02210-2600 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
12 March 2009

WO 2008/141174 A3

(54) Title: GENE KNOCKOUT MESOPHILIC AND THERMOPHILIC ORGANISMS, AND METHODS OF USE THEREOF

(57) Abstract: One aspect of the invention relates to a genetically modified thermophilic or mesophilic microorganism, wherein a first native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which first native gene encodes a first native enzyme involved in the metabolic production of an organic acid or a salt thereof, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product. In certain embodiments, the aforementioned microorganism further comprises a first non-native gene, which first non- native gene encodes a first non-native enzyme involved in the metabolic production of ethanol. Another aspect of the invention relates to a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism.

Gene Knockout Mesophilic and Thermophilic Organisms, and Methods of Use Thereof

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/916,978, filed May 9, 2007; the entire contents of which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

Energy conversion, utilization and access underlie many of the great challenges of our time, including those associated with sustainability, environmental quality, security, and poverty. New applications of emerging technologies are required to respond to these challenges. Biotechnology, one of the most powerful of the emerging technologies, can give rise to important new energy conversion processes. Plant biomass and derivatives thereof are a resource for the biological conversion of energy to forms useful to humanity.

Among forms of plant biomass, lignocellulosic biomass ('biomass') is particularly well-suited for energy applications because of its large-scale availability, low cost, and environmentally benign production. In particular, many energy production and utilization cycles based on cellulosic biomass have near-zero greenhouse gas emissions on a life-cycle basis. The primary obstacle impeding the more widespread production of energy from biomass feedstocks is the general absence of low-cost technology for overcoming the recalcitrance of these materials to conversion into useful fuels. Lignocellulosic biomass contains carbohydrate fractions (*e.g.*, cellulose and hemicellulose) that can be converted into ethanol. In order to convert these fractions, the cellulose and hemicellulose must ultimately be converted or hydrolyzed into monosaccharides; it is the hydrolysis that has historically proven to be problematic.

Biologically mediated processes are promising for energy conversion, in particular for the conversion of lignocellulosic biomass into fuels. Biomass processing schemes involving enzymatic or microbial hydrolysis commonly involve four biologically mediated transformations: (1) the production of saccharolytic enzymes (cellulases and

hemicellulases); (2) the hydrolysis of carbohydrate components present in pretreated biomass to sugars; (3) the fermentation of hexose sugars (*e.g.*, glucose, mannose, and galactose); and (4) the fermentation of pentose sugars (*e.g.*, xylose and arabinose). These four transformations occur in a single step in a process configuration called consolidated bioprocessing (CBP), which is distinguished from other less highly integrated configurations in that it does not involve a dedicated process step for cellulase and/or hemicellulase production.

CBP offers the potential for lower cost and higher efficiency than processes featuring dedicated cellulase production. The benefits result in part from avoided capital costs, substrate and other raw materials, and utilities associated with cellulase production. In addition, several factors support the realization of higher rates of hydrolysis, and hence reduced reactor volume and capital investment using CBP, including enzyme-microbe synergy and the use of thermophilic organisms and/or complexed cellulase systems. Moreover, cellulose-adherent cellulolytic microorganisms are likely to compete successfully for products of cellulose hydrolysis with non-adhered microbes, e.g., contaminants, which could increase the stability of industrial processes based on microbial cellulose utilization. Progress in developing CBP-enabling microorganisms is being made through two strategies: engineering naturally occurring cellulolytic microorganisms to improve product-related properties, such as yield and titer; and engineering non-cellulolytic organisms that exhibit high product yields and titers to express a heterologous cellulase and hemicellulase system enabling cellulose and hemicellulose utilization.

Many bacteria have the ability to ferment simple hexose sugars into a mixture of acidic and pH-neutral products via the process of glycolysis. The glycolytic pathway is abundant and comprises a series of enzymatic steps whereby a six carbon glucose molecule is broken down, via multiple intermediates, into two molecules of the three carbon compound pyruvate. This process results in the net generation of ATP (biological energy supply) and the reduced cofactor NADH.

Pyruvate is an important intermediary compound of metabolism. For example, under aerobic conditions pyruvate may be oxidized to acetyl coenzyme A (acetyl CoA), which then enters the tricarboxylic acid cycle (TCA), which in turn generates synthetic precursors, CO_2 and reduced cofactors. The cofactors are then oxidized by donating hydrogen equivalents, via a series of enzymatic steps, to oxygen resulting in the formation

of water and ATP. This process of energy formation is known as oxidative phosphorylation.

Under anaerobic conditions (no available oxygen), fermentation occurs in which the degradation products of organic compounds serve as hydrogen donors and acceptors. Excess NADH from glycolysis is oxidized in reactions involving the reduction of organic substrates to products, such as lactate and ethanol. In addition, ATP is regenerated from the production of organic acids, such as acetate, in a process known as substrate level phosphorylation. Therefore, the fermentation products of glycolysis and pyruvate metabolism include a variety of organic acids, alcohols and CO_2.

The majority of facultative anaerobic bacteria do not produce high yields of ethanol under either aerobic or anaerobic conditions. Most facultative anaerobes metabolize pyruvate aerobically via pyruvate dehydrogenase (PDH) and the tricarboxylic acid cycle (TCA). Under anaerobic conditions, the main energy pathway for the metabolism of pyruvate is via pyruvate-formate-lyase (PFL) pathway to give formate and acetyl-CoA. Acetyl-CoA is then converted to acetate, via phosphotransacetylase (PTA) and acetate kinase (ACK) with the co-production of ATP, or reduced to ethanol via acetalaldehyde dehydrogenase (AcDH) and alcohol dehydrogenase (ADH). In order to maintain a balance of reducing equivalents, excess NADH produced from glycolysis is re-oxidized to NAD^+ by lactate dehydrogenase (LDH) during the reduction of pyruvate to lactate. NADH can also be re-oxidized by AcDH and ADH during the reduction of acetyl-CoA to ethanol, but this is a minor reaction in cells with a functional LDH. Theoretical yields of ethanol are therefore not achieved since most acetyl CoA is converted to acetate to regenerate ATP and excess NADH produced during glycolysis is oxidized by LDH.

Metabolic engineering of microorganisms could also result in the creation of a targeted knockout of the genes encoding for the production of enzymes, such as lactate dehydrogenase. In this case, "knock out" of the genes means partial, substantial, or complete deletion, silencing, inactivation, or down-regulation. If the conversion of pyruvate to lactate (the salt form of lactic acid) by the action of LDH was not available in the early stages of the glycolytic pathway, then the pyruvate could be more efficiently converted to acetyl CoA by the action of pyruvate dehydrogenase or pyruvate-ferredoxin oxidoreductase. If the further conversion of acetyl CoA to acetate (the salt form of acetic acid) by phosphotransacetylase and acetate kinase was also not available, *i.e.*, if the genes

encoding for the production of PTA and ACK were knocked out, then the acetyl CoA could be more efficiently converted to ethanol by AcDH and ADH. Accordingly, a genetically modified strain of microorganism with such targeted gene knockouts, which eliminates the production of organic acids, would have an increased ability to produce ethanol as a fermentation product.

Ethanologenic organisms, such as *Zymomonas mobilis*, *Zymobacter palmae*, *Acetobacter pasteurianus*, or *Sarcina ventriculi*, and some yeasts (*e.g.*, *Saccharomyces cerevisiae*), are capable of a second type of anaerobic fermentation, commonly referred to as alcoholic fermentation, in which pyruvate is metabolized to acetaldehyde and CO_2 by pyruvate decarboxylase (PDC). Acetaldehyde is then reduced to ethanol by ADH regenerating NAD^+. Alcoholic fermentation results in the metabolism of one molecule of glucose to two molecules of ethanol and two molecules of CO_2. If the conversion of pyruvate to undesired organic acids could be avoided, as detailed above, then such a genetically modified microorganism would have an increased ability to produce ethanol as a fermentation product.

Any discussion of the prior art throughout the specification should in no way be considered as an admission that such prior art is widely known or forms part of common general knowledge in the field.

SUMMARY OF THE INVENTION

According to a first aspect, the present invention provides an isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, or a complement thereof.

In certain embodiments, the invention relates to the aforementioned nucleic acid molecule which shares at least about 95% sequence identity to the nucleotide sequence of any one of SEQ ID NOS:1-5, 30-31, and 47-61, or a complement thereof.

According to a second aspect, the present invention provides a genetic construct comprising SEQ ID NO:1 operably linked to a promoter expressible in a thermophilic or mesophilic bacterium.

According to a third aspect, the present invention provides a recombinant thermophilic or mesophilic bacterium comprising the genetic construct of the second aspect.

According to a fourth aspect, the present invention provides a vector comprising the nucleic acid molecule of the first aspect.

According to a fifth aspect, the present invention provides a host cell comprising the nucleic acid molecule of the first aspect.

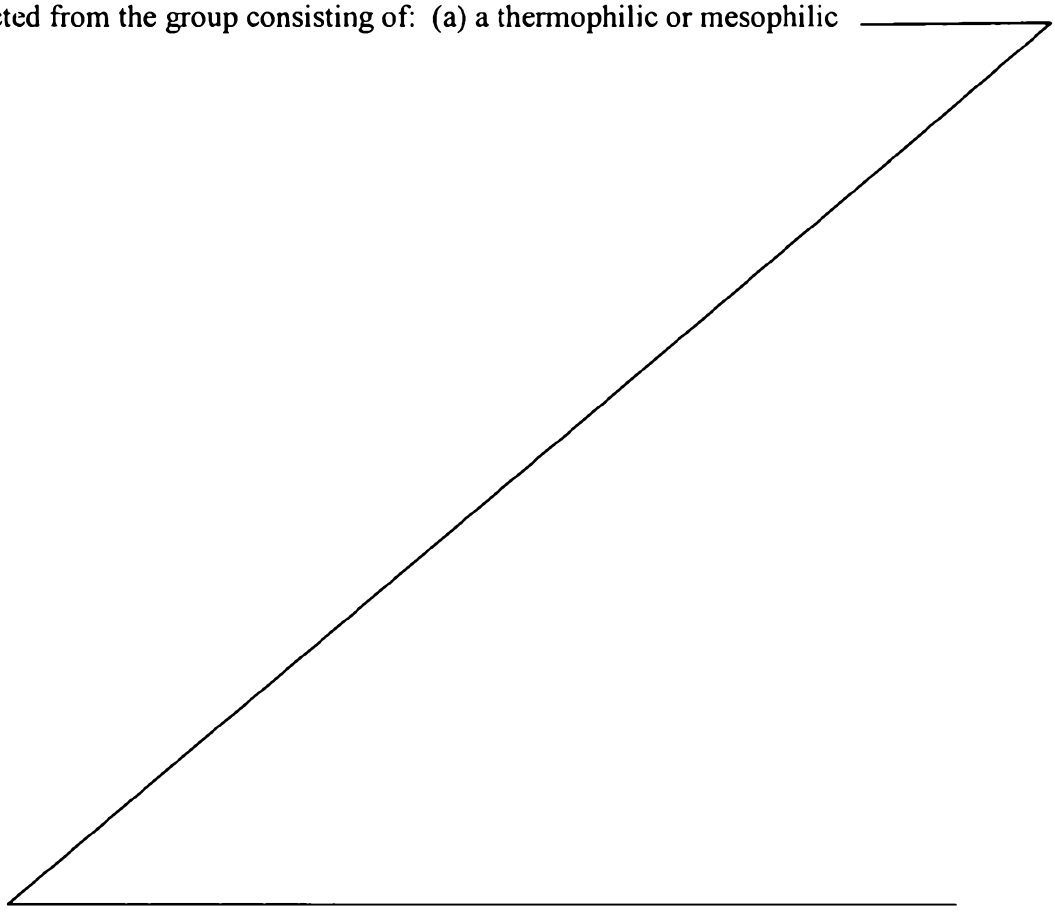
In certain embodiments, the invention relates to the aforementioned host cell, wherein said host cell is a thermophilic or mesophilic bacterial cell.

According to a sixth aspect, the present invention provides a genetically modified thermophilic or mesophilic microorganism, wherein the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a Gram-negative bacterium or a Gram-positive bacterium. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a species of the genera *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, or *Anoxybacillus*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is a bacterium selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium*

2008251465 05 Nov 2012

straminosolvens, Geobacillus thermoglucosidasius, Geobacillus stearothermophilus, Saccharococcus caldoxylosilyticus, Saccharoccus thermophilus, Paenibacillus campinasensis, Bacillus flavothermus, Anoxybacillus kamchatkensis, Anoxybacillus gonensis, Caldicellulosiruptor acetigenus, Caldicellulosiruptor saccharolyticus, Caldicellulosiruptor kristjanssonii, Caldicellulosiruptor owensensis, Caldicellulosiruptor lactoaceticus, and *Anaerocellum thermophilum.* In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Thermoanaerobacterium saccharolyticum.* In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic



microorganism with a native ability to metabolize a hexose sugar; (b) a thermophilic or mesophilic microorganism with a native ability to metabolize a pentose sugar; and (c) a thermophilic or mesophilic microorganism with a native ability to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to metabolize a hexose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Clostridium straminisolvens* or *Clostridium thermocellum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is *Clostridium cellulolyticum*, *Clostridium kristjanssonii*, or *Clostridium stercorarium subsp. leptosaprartum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to metabolize a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is selected from the group consisting of *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium polysaccharolyticum*, and *Thermoanaerobacterium thermosaccharolyticum*. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a hexose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is selected from the group consisting of lactic acid and acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic

acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is lactate dehydrogenase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is acetate kinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a second native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is acetate kinase or phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is lactate dehydrogenase.

According to a seventh aspect, the present invention provides a genetically modified thermophilic or mesophilic microorganism, wherein (a) the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1; and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the metabolic production of ethanol, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar. In certain

embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or

mesophilic microorganism to metabolize a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar and a pentose sugar. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH). In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second non-native enzyme is xylose isomerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene corresponds to SEQ ID NOS:6, 10, or 14. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is xylulokinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NOS:7, 11, or 15. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is L-arabinose isomerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NOS:8 or 12. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native enzyme is L-ribulose-5-phosphate 4-epimerase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said non-native gene corresponds to SEQ ID NO:9 or 13. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said

microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose; (b) a thermophilic or mesophilic microorganism with a native ability to hydrolyze xylan; and (c) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose and xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze cellulose and xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism has a native ability to hydrolyze xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is selected from the group consisting of lactic acid and acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is selected from the group consisting of lactate dehydrogenase, acetate kinase, and phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is lactate dehydrogenase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is acetate kinase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first native enzyme is phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein a second native gene is

partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which second native gene encodes a second native enzyme involved in the metabolic production of an organic acid or a salt thereof. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is acetate kinase or phosphotransacetylase. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said second native enzyme is lactate dehydrogenase.

According to an eighth aspect, the present invention provides a genetically modified thermophilic or mesophilic microorganism, wherein (a) the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1; and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the hydrolysis of a polysaccharide, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose and xylan. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said organic acid is lactic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified

microorganism, wherein said organic acid is acetic acid. In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH). In certain embodiments, the present invention relates to the aforementioned genetically modified microorganism, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol.

In certain embodiments, the present invention relates to any of the aforementioned genetically modified microorganisms, wherein said microorganism is mesophilic. In certain embodiments, the present invention relates to any of the aforementioned genetically modified microorganisms, wherein said microorganism is thermophilic.

According to a ninth aspect, the present invention provides a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to the sixth or seventh aspect.

According to a tenth aspect, the present invention provides a process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to the sixth or seventh aspect.

In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, and combinations thereof. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is selected from the group consisting of corn stover, sugarcane bagasse, switchgrass, and poplar wood. In certain embodiments, the present invention relates to the aforementioned process, wherein said

lignocellulosic biomass is corn stover. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is sugarcane bagasse. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is switchgrass. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is poplar wood. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is willow. In certain embodiments, the present invention relates to the aforementioned process, wherein said lignocellulosic biomass is paper sludge.

According to an eleventh aspect, the present invention provides ethanol produced by the process according to the ninth aspect or tenth aspect.

Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise", "comprising", and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to".

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the glycolysis pathway

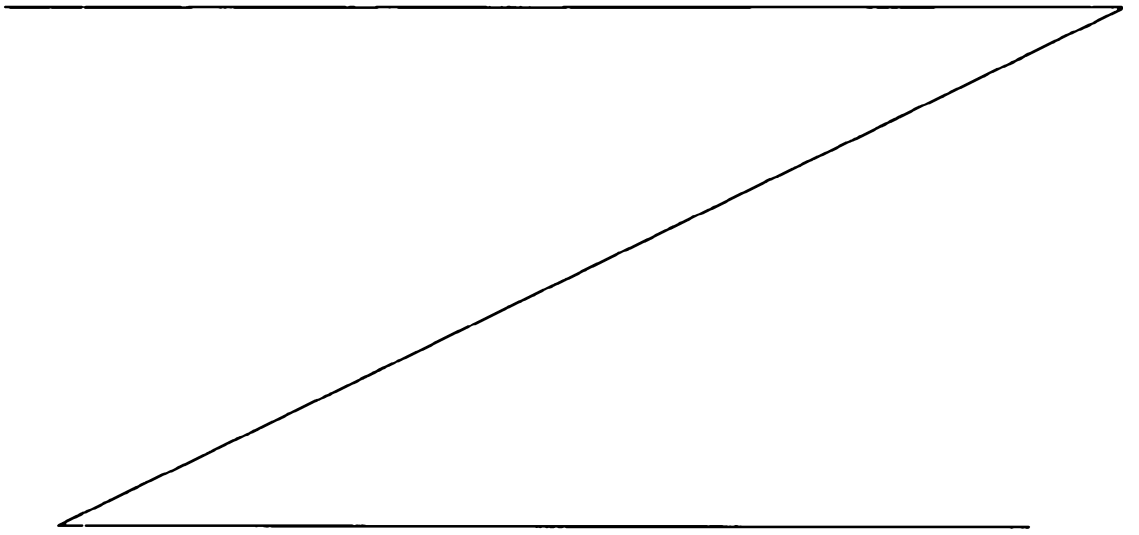


Figure 2 depicts pentose and glucuronate interconversions and highlights the enzymes, xylose isomerase (XI or 5.3.1.5) and xylulokinase (XK or 2.7.1.17), in the D-xylose to ethanol pathway.

Figure 3 depicts pentose and glucuronate interconversions and highlights the enzymes, L-arabinose isomerase (5.3.1.4) and L-ribulose-5-phosphate 4-epimerase (5.1.3.4), in the L-arabinose utilization pathway.

Figure 4 depicts pentose and glucuronate interconversions and shows that the genes for xylose isomerase, xylulokinase, L-arabinose isomerase, and L-ribulose-5-phosphate 4-epimerase are present in *C. cellulolyticum*.

Figure 5 depicts pentose and glucuronate interconversions and shows that xylose isomerase and xylulokinase are present, while L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase are absent in *C. phytofermentans*.

Figure 6 shows an alignment of *Clostridium thermocellum* (SEQ ID NO: 77), *Clostridium cellulolyticum* (SEQ ID NO: 78), *Thermoanaerobacterium saccharolyticum* (SEQ ID NO: 79), *C. stercorarium* (SEQ ID NO: 80), *C. stercorarium II* (SEQ ID NO: 81), *Caldiscellulosiruptor kristjanssonii* (SEQ ID NO: 82), *C. phytofermentans* (SEQ ID NO: 83), indicating about 73-89% homology at the level of the 16S rDNA gene.

Figure 7 shows the construction of a double crossover knockout vector for inactivation of the *ack* gene in *Clostridium thermocellum* based on the plasmid pIKM1.

Figure 8 shows the construction of a double crossover knockout vector for inactivation of the *ack* gene in *Clostridium thermocellum* based on the replicative plasmid pNW33N.

Figure 9 shows the construction of a double crossover knockout vector for inactivation of the *ldh* gene in *Clostridium thermocellum* based on the plasmid pIKM1.

Figure 10 shows the construction of a double crossover knockout vector for inactivation of the *ldh* gene in *Clostridium thermocellum* based on the replicative plasmid vector pNW33N.

Figure 11 shows the construction of a double crossover suicide vector for inactivation of the ldh gene in *Clostridium thermocellum* based on the plasmid pUC19.

Figures 12A and **12B** show product formation and OD_{600} for *C. straminisolvens* grown on cellobiose and Avicel®, respectively.

Figures 13A and **13B** show product formation and OD_{600} for *C. thermocellum* grown on cellobiose and Avicel®, respectively.

Figures 14A and **14B** show product formation and OD_{600} for *C. cellulolyticum* grown on cellobiose and Avicel®, respectively.

Figures 15A and **15B** show product formation and OD_{600} for *C. stercorarium* subs. *leptospartum* grown on cellobiose and Avicel®, respectively.

Figures 16A and **16B** show product formation and OD_{600} for *Caldicellulosiruptor kristjanssonii* grown on cellobiose and Avicel®, respectively.

Figures 17A and **17B** show product formation and OD_{600} for *Clostridium phytofermentans* grown on cellobiose and Avicel®, respectively.

Figure 18 shows total metabolic byproducts after 48 hours of fermentation of 2.5 g/L xylan and 2.5 g/L cellobiose.

Figure 19 shows a map of the *ack* gene and the region amplified by PCR for gene disruption.

Figure 20 shows a map of the *ldh* 2262 gene and the region amplified by PCR for gene disruption.

Figure 21 shows an example of *C. cellulolyticum (C. cell.) ldh* (2262) double crossover knockout fragment.

Figure 22 shows a map of the *ack* gene of *Clostridium phytofermentans* and the region amplified by PCR for gene disruption.

Figure 23 shows an example of a putative double crossover knockout construct with the mL*s* gene as a selectable marker in *Clostridium phytofermentans*.

Figure 24 shows a map of the *ldh* 1389 gene and the region amplified by PCR for gene disruption.

Figure 25 shows an example of a putative double crossover knockout construct with the mL*s* gene as a selectable marker.

Figure 26 is a diagram representing bp 250-550 of pMOD™-2<MCS> (SEQ ID NO: 84).

Figure 27 shows the product concentration profiles for 1% Avicel® using *C. straminisolvens*. The ethanol-to-acetate ratio is depicted as E/A and the ratio of ethanol-to-total products is depicted as E/T.

Figure 28 shows an example of a vector for retargeting the Ll.LtrB intron to insert in *C. cell.* ACK gene (SEQ ID NO:21).

Figure 29 shows an example of vector for retargeting the Ll.LtrB intron to insert in *C. cell.* LDH2744 gene (SEQ ID NO:23).

Figure 30 shows an alignment of *T. pseudoethanolicus* 39E (SEQ ID NO: 85), *T. sp* strain 59 (SEQ ID NO: 86), *T. saccharolyticum* B6A-RI (SEQ ID NO: 87), *T. saccharolyticum* YS485 (SEQ ID NO: 88) and consensus (SEQ ID NO: 89) at the level of the 16S rDNA gene.

Figure 31 shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 36), *T. pseudoethanolicus* (SEQ ID NO: 35), *T. saccharolyticum* B6A-RI (SEQ ID NO: 38), *T. saccharolyticum* YS485 (SEQ ID NO: 32) and consensus (SEQ ID NO: 90) at the level of the pta gene.

Figure 32 shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 37), *T. pseudoethanolicus* (SEQ ID NO: 34), *T. saccharolyticum* B6A-RI (SEQ ID NO: 39), *T. saccharolyticum* YS485 (SEQ ID NO: 33) and consensus (SEQ ID NO: 91) at the level of the ack gene.

Figure 33 shows an alignment of *T. sp.* strain 59 (SEQ ID NO: 41), *T. pseudoethanolicus* 39E (SEQ ID NO: 42), *T. saccharolyticum* B6A-RI (SEQ ID NO: 43), *T. saccharolyticum* YS485 (SEQ ID NO: 40) and consensus (SEQ ID NO: 92) at the level of the ldh gene.

Figure 34 shows a schematic of the glycolysis/fermentation pathway.

Figure 35 shows an example of a pMU340 plasmid.

Figure 36 shows an example of a pMU102 *Z. mobilis* PDC-ADH plasmid.

Figure 37 shows an example of a pMU102 *Z. palmae* PDC, *Z. mobilis* ADH plasmid.

Figure 38 shows the plasmid map of pMU360. The DNA sequence of pMU360 is set forth as SEQ ID NO:61.

Figure 39 shows the lactate levels in nine colonies of thiamphenicol-resistant transformants.

Figure 40 shows an example of a *T. sacch.* pfl KO single crossover plasmid (SEQ ID NO:47).

Figure 41 shows an example of a *T. sacch.* pfl KO double crossover plasmid (SEQ ID NO:48).

Figure 42 shows an example of a *C. therm.* pfl KO single crossover plasmid (SEQ ID NO:49).

Figure 43 shows an example of a *C. therm.* pfl KO double crossover plasmid (SEQ ID NO:50).

Figure 44 shows an example of a *C. phyto.* pfl KO single crossover plasmid (SEQ ID NO:51).

Figure 45 shows an example of a *C. phyto.* pfl KO double crossover plasmid (SEQ ID NO:52).

Figure 46 shows an example of a *T. sacch.* #59 L-ldh KO single crossover plasmid (SEQ ID NO:53).

Figure 47 shows an example of a *T. sacch.* #59 L-ldh KO double crossover plasmid (SEQ ID NO:54).

Figure 48 shows an example of a *T. sacch.* #59 pta/ack KO single crossover plasmid (SEQ ID NO:55).

Figure 49 shows an example of a *T. sacch.* #59 pta/ack KO double crossover plasmid (SEQ ID NO:56).

Figure 50 shows an example of a *T. pseudo.* L-ldh KO single crossover plasmid (SEQ ID NO:57).

Figure 51 shows an example of a *T. pseudo.* L-ldh KO double crossover plasmid (SEQ ID NO:58).

Figure 52 shows an example of a *T. pseudo.* ack KO single crossover plasmid (SEQ ID NO:59).

Figure 53 shows an example of a *T. pseudo.* pta/ack KO double crossover plasmid (SEQ ID NO:60).

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes representative highly cellulolytic organisms.

Table 2 summarizes representative native cellulolytic and xylanolytic organisms.

Table 3 shows a categorization of bacterial strains based on their substrate utilization.

Table 4 shows insertion location and primers to retarget Intron to *C. cellulolyticum* acetate kinase.

Table 5 shows insertion location and primers to retarget Intron to *C. cellulolyticum* lactate dehydrogenase.

Table 6 shows fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains.

DETAILED DESCRIPTION OF THE INVENTION

Aspects of the present invention relate to the engineering of thermophilic or mesophilic microorganisms for use in the production of ethanol from lignocellulosic biomass. The use of thermophilic bacteria for ethanol production offers many advantages over traditional processes based upon mesophilic ethanol producers. For example, the use

of thermophilic organisms provides significant economic savings over traditional process methods due to lower ethanol separation costs, reduced requirements for external enzyme addition, and reduced processing times.

Aspects of the present invention relate to a process by which the cost of ethanol production from cellulosic biomass-containing materials can be reduced by using a novel processing configuration. In particular, the present invention provides numerous methods for increasing ethanol production in a genetically modified microorganism.

In certain other embodiments, the present invention relates to genetically modified thermophilic or mesophilic microorganisms, wherein a gene or a particular polynucleotide sequence is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which gene or polynucleotide sequence encodes for an enzyme that confers upon the microorganism the ability to produce organic acids as fermentation products, thereby increasing the ability of the microorganism to produce ethanol as the major fermentation product. Further, by virtue of a novel integration of processing steps, commonly known as consolidated bioprocessing, aspects of the present invention provide for more efficient production of ethanol from cellulosic-biomass-containing raw materials. The incorporation of genetically modified thermophilic or mesophilic microorganisms in the processing of said materials allows for fermentation steps to be conducted at higher temperatures, improving process economics. For example, reaction kinetics are typically proportional to temperature, so higher temperatures are generally associated with increases in the overall rate of production. Additionally, higher temperature facilitates the removal of volatile products from the broth and reduces the need for cooling after pretreatment.

In certain embodiments, the present invention relates to genetically modified or recombinant thermophilic or mesophilic microorganisms with increased ability to produce enzymes that confer the ability to produce ethanol as a fermentation product, the presence of which enzyme(s) modify the process of metabolizing lignocellulosic biomass materials to produce ethanol as the major fermentation product. In one aspect of the invention, one or more non-native genes are inserted into a genetically modified thermophilic or mesophilic microorganism, wherein said non-native gene encodes an enzyme involved in the metabolic production of ethanol, for example, such enzyme may confer the ability to metabolize a pentose sugar and/or a hexose sugar. For example, in one embodiment, the enzyme may be involved in the D-xylose or L-arabinose pathway, thereby allowing the microorganism to

metabolize a pentose sugar, *i.e.*, D-xylose or L-arabinose. By inserting (*e.g.*, introducing or adding) a non-native gene that encodes an enzyme involved in the metabolism or utilization of D-xylose or L-arabinose, the microorganism has an increased ability to produce ethanol relative to the native organism.

The present invention also provides novel compositions that may be integrated into the microorganisms of the invention. In one embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of a nucleotide sequence shown in any one of SEQ ID NOS:1-76. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, is one which is sufficiently complementary to a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, such that it can hybridize to a nucleotide sequence shown in any one of SEQ ID NOS:1-76, or the coding region thereof, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the nucleotide sequences (*e.g.*, to the entire length of the nucleotide sequence) shown in any one of SEQ ID NOS:1-76, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecules of the invention may comprise only a portion of the nucleic acid sequence of any one of SEQ ID NOS:1-76, or the coding region thereof; for example, the nucleic acid molecule may be a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a protein. In another embodiment, the nucleic acid molecules may comprise at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of any one of SEQ ID NOS:1-76.

Definitions

The term "heterologous polynucleotide segment" is intended to include a polynucleotide segment that encodes one or more polypeptides or portions or fragments of polypeptides. A heterologous polynucleotide segment may be derived from any source, *e.g.*, eukaryotes, prokaryotes, viruses, or synthetic polynucleotide fragments.

The terms "promoter" or "surrogate promoter" is intended to include a polynucleotide segment that can transcriptionally control a gene-of-interest that it does not transcriptionally control in nature. In certain embodiments, the transcriptional control of a surrogate promoter results in an increase in expression of the gene-of-interest. In certain embodiments, a surrogate promoter is placed 5' to the gene-of-interest. A surrogate promoter may be used to replace the natural promoter, or may be used in addition to the natural promoter. A surrogate promoter may be endogenous with regard to the host cell in which it is used, or it may be a heterologous polynucleotide sequence introduced into the host cell, *e.g.*, exogenous with regard to the host cell in which it is used.

The terms "gene(s)" or "polynucleotide segment" or "polynucleotide sequence(s)" are intended to include nucleic acid molecules, *e.g.*, polynucleotides which include an open reading frame encoding a polypeptide, and can further include non-coding regulatory sequences, and introns. In addition, the terms are intended to include one or more genes that map to a functional locus. In addition, the terms are intended to include a specific gene for a selected purpose. The gene may be endogenous to the host cell or may be recombinantly introduced into the host cell, *e.g.*, as a plasmid maintained episomally or a plasmid (or fragment thereof) that is stably integrated into the genome. In addition to the plasmid form, a gene may, for example, be in the form of linear DNA. In certain embodiments, the gene of polynucleotide segment is involved in at least one step in the bioconversion of a carbohydrate to ethanol. Accordingly, the term is intended to include any gene encoding a polypeptide, such as the enzymes acetate kinase (ACK), phosphotransacetylase (PTA), and/or lactate dehydrogenase (LDH), enzymes in the D-xylose pathway, such as xylose isomerase and xylulokinase, enzymes in the L-arabinose pathway, such as L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase. The term gene is also intended to cover all copies of a particular gene, *e.g.*, all of the DNA sequences in a cell encoding a particular gene product.

The term "transcriptional control" is intended to include the ability to modulate gene expression at the level of transcription. In certain embodiments, transcription, and thus

gene expression, is modulated by replacing or adding a surrogate promoter near the 5' end of the coding region of a gene-of-interest, thereby resulting in altered gene expression. In certain embodiments, the transcriptional control of one or more gene is engineered to result in the optimal expression of such genes, *e.g.*, in a desired ratio. The term also includes inducible transcriptional control as recognized in the art.

The term "expression" is intended to include the expression of a gene at least at the level of mRNA production.

The term "expression product" is intended to include the resultant product, *e.g.*, a polypeptide, of an expressed gene.

The term "increased expression" is intended to include an alteration in gene expression at least at the level of increased mRNA production and, preferably, at the level of polypeptide expression. The term "increased production" is intended to include an increase in the amount of a polypeptide expressed, in the level of the enzymatic activity of the polypeptide, or a combination thereof.

The terms "activity," "activities," "enzymatic activity," and "enzymatic activities" are used interchangeably and are intended to include any functional activity normally attributed to a selected polypeptide when produced under favorable conditions. Typically, the activity of a selected polypeptide encompasses the total enzymatic activity associated with the produced polypeptide. The polypeptide produced by a host cell and having enzymatic activity may be located in the intracellular space of the cell, cell-associated, secreted into the extracellular milieu, or a combination thereof. Techniques for determining total activity as compared to secreted activity are described herein and are known in the art.

The term "xylanolytic activity" is intended to include the ability to hydrolyze glycosidic linkages in oligopentoses and polypentoses.

The term "cellulolytic activity" is intended to include the ability to hydrolyze glycosidic linkages in oligohexoses and polyhexoses. Cellulolytic activity may also include the ability to depolymerize or debranch cellulose and hemicellulose.

As used herein, the term "lactate dehydrogenase" or "LDH" is intended to include the enzyme capable of converting pyruvate into lactate. It is understood that LDH can also catalyze the oxidation of hydroxybutyrate.

As used herein the term "alcohol dehydrogenase" or "ADH" is intended to include the enzyme capable of converting acetaldehyde into an alcohol, advantageously, ethanol.

The term "pyruvate decarboxylase activity" is intended to include the ability of a polypeptide to enzymatically convert pyruvate into acetaldehyde (*e.g.*, "pyruvate decarboxylase" or "PDC"). Typically, the activity of a selected polypeptide encompasses the total enzymatic activity associated with the produced polypeptide, comprising, *e.g.*, the superior substrate affinity of the enzyme, thermostability, stability at different pHs, or a combination of these attributes.

The term "ethanologenic" is intended to include the ability of a microorganism to produce ethanol from a carbohydrate as a fermentation product. The term is intended to include, but is not limited to, naturally occurring ethanologenic organisms, ethanologenic organisms with naturally occurring or induced mutations, and ethanologenic organisms which have been genetically modified.

The terms "fermenting" and "fermentation" are intended to include the enzymatic process (*e.g.*, cellular or acellular, *e.g.*, a lysate or purified polypeptide mixture) by which ethanol is produced from a carbohydrate, in particular, as a product of fermentation.

The term "secreted" is intended to include the movement of polypeptides to the periplasmic space or extracellular milieu. The term "increased secretion" is intended to include situations in which a given polypeptide is secreted at an increased level (*i.e.*, in excess of the naturally-occurring amount of secretion). In certain embodiments, the term "increased secreted" refers to an increase in secretion of a given polypeptide that is at least about 10% or at least about 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, as compared to the naturally-occurring level of secretion.

The term "secretory polypeptide" is intended to include any polypeptide(s), alone or in combination with other polypeptides, that facilitate the transport of another polypeptide from the intracellular space of a cell to the extracellular milieu. In certain embodiments, the secretory polypeptide(s) encompass all the necessary secretory polypeptides sufficient

to impart secretory activity to a Gram-negative or Gram-positive host cell. Typically, secretory proteins are encoded in a single region or locus that may be isolated from one host cell and transferred to another host cell using genetic engineering. In certain embodiments, the secretory polypeptide(s) are derived from any bacterial cell having secretory activity. In certain embodiments, the secretory polypeptide(s) are derived from a host cell having Type II secretory activity. In certain embodiments, the host cell is a thermophilic bacterial cell.

The term "derived from" is intended to include the isolation (in whole or in part) of a polynucleotide segment from an indicated source or the purification of a polypeptide from an indicated source. The term is intended to include, for example, direct cloning, PCR amplification, or artificial synthesis from or based on a sequence associated with the indicated polynucleotide source.

By "thermophilic" is meant an organism that thrives at a temperature of about 45°C or higher.

By "mesophilic" is meant an organism that thrives at a temperature of about 20-45°C.

The term "organic acid" is art-recognized. The term "lactic acid" refers to the organic acid 2-hydroxypropionic acid in either the free acid or salt form. The salt form of lactic acid is referred to as "lactate" regardless of the neutralizing agent, *i.e.*, calcium carbonate or ammonium hydroxide. The term "acetic acid" refers to the organic acid methanecarboxylic acid, also known as ethanoic acid, in either free acid or salt form. The salt form of acetic acid is referred to as "acetate."

Certain embodiments of the present invention provide for the "insertion," (*e.g.*, the addition, integration, incorporation, or introduction) of certain genes or particular polynucleotide sequences within thermophilic or mesophilic microorganisms, which insertion of genes or particular polynucleotide sequences may be understood to encompass "genetic modification(s)" or "transformation(s)" such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be "genetically modified" or "transformed." In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

Certain embodiments of the present invention provide for the "inactivation" or "deletion" of certain genes or particular polynucleotide sequences within thermophilic or mesophilic microorganisms, which "inactivation" or "deletion" of genes or particular polynucleotide sequences may be understood to encompass "genetic modification(s)" or "transformation(s)" such that the resulting strains of said thermophilic or mesophilic microorganisms may be understood to be "genetically modified" or "transformed." In certain embodiments, strains may be of bacterial, fungal, or yeast origin.

The term "CBP organism" is intended to include microorganisms of the invention, *e.g.*, microorganisms that have properties suitable for CBP.

In one aspect of the invention, the genes or particular polynucleotide sequences are inserted to activate the activity for which they encode, such as the expression of an enzyme. In certain embodiments, genes encoding enzymes in the metabolic production of ethanol, *e.g.*, enzymes that metabolize pentose and/or hexose sugars, may be added to a mesophilic or thermophilic organism. In certain embodiments of the invention, the enzyme may confer the ability to metabolize a pentose sugar and be involved, for example, in the D-xylose pathway and/or L-arabinose pathway.

In one aspect of the invention, the genes or particular polynucleotide sequences are partially, substantially, or completely deleted, silenced, inactivated, or down-regulated in order to inactivate the activity for which they encode, such as the expression of an enzyme. Deletions provide maximum stability because there is no opportunity for a reverse mutation to restore function. Alternatively, genes can be partially, substantially, or completely deleted, silenced, inactivated, or down-regulated by insertion of nucleic acid sequences that disrupt the function and/or expression of the gene (*e.g.*, P1 transduction or other methods known in the art). The terms "eliminate," "elimination," and "knockout" are used interchangeably with the term "deletion." In certain embodiments, strains of thermophilic or mesophilic microorganisms of interest may be engineered by site directed homologous recombination to knockout the production of organic acids. In still other embodiments, RNAi or antisense DNA (asDNA) may be used to partially, substantially, or completely silence, inactivate, or down-regulate a particular gene of interest.

In certain embodiments, the genes targeted for deletion or inactivation as described herein may be endogenous to the native strain of the microorganism, and may thus be

understood to be referred to as "native gene(s)" or "endogenous gene(s)." An organism is in "a native state" if it has not been genetically engineered or otherwise manipulated by the hand of man in a manner that intentionally alters the genetic and/or phenotypic constitution of the organism. For example, wild-type organisms may be considered to be in a native state. In other embodiments, the gene(s) targeted for deletion or inactivation may be non-native to the organism.

Biomass

The terms "lignocellulosic material," "lignocellulosic substrate," and "cellulosic biomass" mean any type of biomass comprising cellulose, hemicellulose, lignin, or combinations thereof, such as but not limited to woody biomass, forage grasses, herbaceous energy crops, non-woody-plant biomass, agricultural wastes and/or agricultural residues, forestry residues and/or forestry wastes, paper-production sludge and/or waste paper sludge, waste-water-treatment sludge, municipal solid waste, corn fiber from wet and dry mill corn ethanol plants, and sugar-processing residues.

In a non-limiting example, the lignocellulosic material can include, but is not limited to, woody biomass, such as recycled wood pulp fiber, sawdust, hardwood, softwood, and combinations thereof; grasses, such as switch grass, cord grass, rye grass, reed canary grass, miscanthus, or a combination thereof; sugar-processing residues, such as but not limited to sugar cane bagasse; agricultural wastes, such as but not limited to rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, and corn fiber; stover, such as but not limited to soybean stover, corn stover; and forestry wastes, such as but not limited to recycled wood pulp fiber, sawdust, hardwood (*e.g.*, poplar, oak, maple, birch, willow), softwood, or any combination thereof. Lignocellulosic material may comprise one species of fiber; alternatively, lignocellulosic material may comprise a mixture of fibers that originate from different lignocellulosic materials. Particularly advantageous lignocellulosic materials are agricultural wastes, such as cereal straws, including wheat straw, barley straw, canola straw and oat straw; corn fiber; stovers, such as corn stover and soybean stover; grasses, such as switch grass, reed canary grass, cord grass, and miscanthus; or combinations thereof.

Paper sludge is also a viable feedstock for ethanol production. Paper sludge is solid residue arising from pulping and paper-making, and is typically removed from process

wastewater in a primary clarifier. At a disposal cost of $30/wet ton, the cost of sludge disposal equates to $5/ton of paper that is produced for sale. The cost of disposing of wet sludge is a significant incentive to convert the material for other uses, such as conversion to ethanol. Processes provided by the present invention are widely applicable. Moreover, the saccharification and/or fermentation products may be used to produce ethanol or higher value added chemicals, such as organic acids, aromatics, esters, acetone and polymer intermediates.

Pyruvate formate lyase (PFL)

Pyruvate formate lyase (PFL) is an important enzyme (found in *Escherichia coli* and other organisms) that helps regulate anaerobic glucose metabolism. Using radical chemistry, it catalyzes the reversible conversion of pyruvate and coenzyme-A into formate and acetyl-CoA, a precursor of ethanol. Pyruvate formate lyase is a homodimer made of 85 kDa, 759-residue subunits. It has a 10-stranded beta/alpha barrel motif into which is inserted a beta finger that contains major catalytic residues. The active site of the enzyme, elucidated by x-ray crystallography, holds three essential amino acids that perform catalysis (Gly734, Cys418, and Cys419), three major residues that hold the substrate pyruvate close by (Arg435, Arg176, and Ala272), and two flanking hydrophobic residues (Trp333 and Phe432).

Studies have found structural similarities between the active site of pyruvate formate lyase and that of Class I and Class III ribonucleotide reductase (RNR) enzymes. The roles of the 3 catalytic residues are as follows: Gly734 (glycyl radical) – transfers the radical on and off Cys418, via Cys419; Cys418 (thiyl radical) – performs acylation chemistry on the carbon atom of the pyruvate carbonyl; Cys419 (thiyl radical) – performs hydrogen-atom transfers.

The proposed mechanism for pyruvate formate lyase begins with radical transfer from Gly734 to Cys418, via Cys419. The Cys418 thiyl radical adds covalently to C2 (second carbon atom) of pyruvate, generating an acetyl-enzyme intermediate (which now contains the radical). The acetyl-enzyme intermediate releases a formyl radical that undergoes hydrogen-atom transfer with Cys419. This generates formate and a Cys419 radical. Coenzyme-A undergoes hydrogen-atom transfer with the Cys419 radical to generate a coenzyme-A radical. The coenzyme-A radical then picks up the acetyl group

from Cys418 to generate acetyl-CoA, leaving behind a Cys418 radical. Pyruvate formate lyase can then undergo radical transfer to put the radical back onto Gly734. Each of the above mentioned steps are also reversible.

Two additional enzymes regulate the "on" and "off" states of pyruvate formate lyase to regulate anaerobic glucose metabolism: PFL activase (AE) and PFL deactivase (DA). Activated pyruvate formate lyase allows formation of acetyl-CoA, a small molecule important in the production of energy, when pyruvate is available. Deactivated pyruvate formate lyase, even with substrates present, does not catalyze the reaction. PFL activase is part of the radical SAM (S-adenosylmethionine) superfamily.

The enzyme turns pyruvate formate lyase "on" by converting Gly734 (G-H) into a Gly734 radical (G*) via a 5'-deoxyadenosyl radical (radical SAM). PFL deactivase (DA) turns pyruvate formate lyase "off" by quenching the Gly734 radical. Furthermore, pyruvate formate lyase is sensitive to molecular oxygen (O_2), the presence of which shuts the enzyme off.

Xylose metabolism

Xylose is a five-carbon monosaccharide that can be metabolized into useful products by a variety of organisms. There are two main pathways of xylose metabolism, each unique in the characteristic enzymes they utilize. One pathway is called the "Xylose Reductase-Xylitol Dehydrogenase" or XR-XDH pathway. Xylose reductase (XR) and xylitol dehydrogenase (XDH) are the two main enzymes used in this method of xylose degradation. XR, encoded by the XYL1 gene, is responsible for the reduction of xylose to xylitol and is aided by cofactors NADH or NADPH. Xylitol is then oxidized to xylulose by XDH, which is expressed through the XYL2 gene, and accomplished exclusively with the cofactor NAD+. Because of the varying cofactors needed in this pathway and the degree to which they are available for usage, an imbalance can result in an overproduction of xylitol byproduct and an inefficient production of desirable ethanol. Varying expression of the XR and XDH enzyme levels have been tested in the laboratory in the attempt to optimize the efficiency of the xylose metabolism pathway.

The other pathway for xylose metabolism is called the "Xylose Isomerase" (XI) pathway. Enzyme XI is responsible for direct conversion of xylose into xylulose, and does not proceed via a xylitol intermediate. Both pathways create xylulose, although the

enzymes utilized are different. After production of xylulose both the XR-XDH and XI pathways proceed through enzyme xylulokinase (XK), encoded on gene XKS1, to further modify xylulose into xylulose-5-P where it then enters the pentose phosphate pathway for further catabolism.

Studies on flux through the pentose phosphate pathway during xylose metabolism have revealed that limiting the speed of this step may be beneficial to the efficiency of fermentation to ethanol. Modifications to this flux that may improve ethanol production include a) lowering phosphoglucose isomerase activity, b) deleting the GND1 gene, and c) deleting the ZWF1 gene (Jeppsson et al., 2002). Since the pentose phosphate pathway produces additional NADPH during metabolism, limiting this step will help to correct the already evident imbalance between NAD(P)H and NAD+ cofactors and reduce xylitol byproduct. Another experiment comparing the two xylose metabolizing pathways revealed that the XI pathway was best able to metabolize xylose to produce the greatest ethanol yield, while the XR-XDH pathway reached a much faster rate of ethanol production (Karhumaa et al., 2007).

Microorganisms

The present invention includes multiple strategies for the development of microorganisms with the combination of substrate-utilization and product-formation properties required for CBP. The "native cellulolytic strategy" involves engineering naturally occurring cellulolytic microorganisms to improve product-related properties, such as yield and titer. The "recombinant cellulolytic strategy" involves engineering natively non-cellulolytic organisms that exhibit high product yields and titers to express a heterologous cellulase system that enables cellulose utilization or hemicellulose utilization or both.

Cellulolytic Microorganisms

Several microorganisms reported in the literature to be cellulolytic or have cellulolytic activity have been characterized by a variety of means, including their ability to grow on microcrystalline cellulose as well as a variety of other sugars. Additionally, the organisms may be characterized by other means, including but not limited to, their ability to depolymerize and debranch cellulose and hemicellulose. *Clostridium thermocellum* (strain DSMZ 1237) was used to benchmark the organisms of interest. As used herein, *C.*

thermocellum may include various strains, including, but not limited to, DSMZ 1237, DSMZ 1313, DSMZ 2360, DSMZ 4150, DSMZ 7072, and ATCC 31924. In certain embodiments of the invention, the strain of *C. thermocellum* may include, but is not limited to, DSMZ 1313 or DSMZ 1237. In another embodiment, particularly suitable organisms of interest for use in the present invention include cellulolytic microorganisms with a greater than 70% 16S rDNA homology to *C. thermocellum*. Alignment of *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Thermoanaerobacterium saccharolyticum*, *C. stercorarium*, *C. stercorarium II*, *Caldiscellulosiruptor kristjanssonii*, *C. phytofermentans* indicate a 73 – 85% homology at the level of the 16S rDNA gene (Figure 6).

Clostridium straminisolvens has been determined to grow nearly as well as *C. thermocellum* on Avicel®. Table 1 summarizes certain highly cellulolytic organisms.

Table 1

Strain	DSMZ No.	T optimum; or range	pH optimum; or range	Gram Stain	Aero-tolerant	Utilizes	Products
Clostridium thermocellum	1313	55-60	7	positive	No	cellobiose, cellulose	acetic acid, lactic acid, ethanol, H_2, CO_2
Clostridium straminisolvens	16021	50-55; 45-60	6.5-6.8; 6.0-8.5	positive	Yes	cellobiose, cellulose	acetic acid, lactic acid, ethanol, H_2, CO_2

Organisms were grown on 20 g/L cellobiose or 20 g/L Avicel®. *C. thermocellum* was grown at 60°C and *C. straminisolvens* was grown at 55°C. Both were pre-cultured from -80°C freezer stock (origin DSMZ) on M122 with 50mM MOPS. During mid to late log growth phase pre-cultures were used to inoculate the batch cultures in 100 mL serum bottles to a working volume of 50 mL. Liquid samples were removed periodically for HPLC analysis of metabolic byproducts and sugar consumption. OD_{600} was taken at each of these time points. Figures 12A and 12B show product formation and OD_{600} for *C. straminisolvens* on cellobiose and Avicel®, respectively. Substantial cellobiose (37%) was consumed with 48 hours before OD dropped and product formation leveled off. Figures 13A and 13B show product

formation and OD_{600} for *C. thermocellum* on cellobiose and Avicel®, respectively. *C. thermocellum* consumed ~60% of cellobiose within 48 hours, at which point product formation leveled out. Inhibition due to formation of organic acids caused incomplete utilization of substrates.

Certain microorganisms, including, for example, *C. thermocellum* and *C. straminisolvens*, cannot metabolize pentose sugars, such as D-xylose or L-arabinose, but are able to metabolize hexose sugars. Both D-xylose and L-arabinose are abundant sugars in biomass with D-xylose accounting for approximately 16 – 20% in soft and hard woods and L-arabinose accounting for approximately 25% in corn fiber. Accordingly, the invention provides genetically-modified cellulolytic microorganisms, with the ability to metabolize pentose sugars, such as D-xylose and L-arabinose, thereby to enhance their use as biocatalysts for fermentation in the biomass-to-ethanol industry.

Cellulolytic and Xylanolytic Microorganisms

Several microorganisms determined from literature to be both cellulolytic and xylanolytic have been characterized by their ability to grow on microcrystalline cellulose and birchwood xylan as well as a variety of other sugars. *Clostridium thermocellum* was used to benchmark the organisms of interest. Of the strains selected for characterization *Clostridium cellulolyticum, Clostridium stercorarium* subs. *leptospartum, Caldicellulosiruptor kristjanssonii* and *Clostridium phytofermentans* grew weakly on Avicel® and well on birchwood xylan. Table 2 summarizes some of the native cellulolytic and xylanolytic organisms.

Table 2

Strain	Source/ No.	T optimum; or range	pH optimum; or range	Gram Stain	Aero-tolerant	Utilizes	Products
Clostridium cellulolyticum	DSM 5812	34	7.2	negative	no	Cellulose, xylan, arabinose, mannose, galactose, xylose, glucose, cellobiose	acetic acid, lactic acid, ethanol, H_2, CO_2

Clostridium stercorarium subs. *leptospartum*	DSM 9219	60-65	7.0-7.5	negative	no	Cellulose, cellobiose, lactose, xylose, melibiose, raffinose, ribose, fructose, sucrose	acetic acid, lactic acid, ethanol, H_2, CO_2
Caldicellulosiruptor kristjanssonii	DSM 12137	78; 45-82	7; 5.8-8.0	negative	No	cellobiose, glucose, xylose, galactose, mannose, cellulose	acetic acid, H_2, CO_2, lactic acid, ethanol formate
Clostridium phytofermentans	ATCC 700394	37; 5 - 45	8.5; 6 - 9	Negative (gram type positive)	no	Cellulose, xylan, cellobiose, fructose, galactose, glucose, lactose, maltose, mannose, ribose, xylose	acetic acid, H_2, CO_2, lactic acid, ethanol formate

Organisms were grown on 20 g/L cellobiose, 20 g/L Avicel® or 5 g/L birchwood xylan. *C. cellulolyticum* was grown at 37°C, *C. stercorarium* subs. *leptospartum* was grown at 60°C, *Caldicellulosiruptor kristjanssonii* was grown at 75°C and *Clostridium phytofermentans* was grown at 37°C. All were pre-cultured from -80°C freezer stock in M122c supplemented with 50mM MOPS. During mid to late log growth phase pre-cultures were used to inoculate the batch cultures in 100 mL serum bottles to a working volume of 50 mL. Liquid samples were removed periodically for HPLC analysis of metabolic byproducts and sugar consumption. OD_{600} was taken at each of these time points. Figures 14A-17B show product formation and OD_{600} for growth on cellobiose and Avicel®.

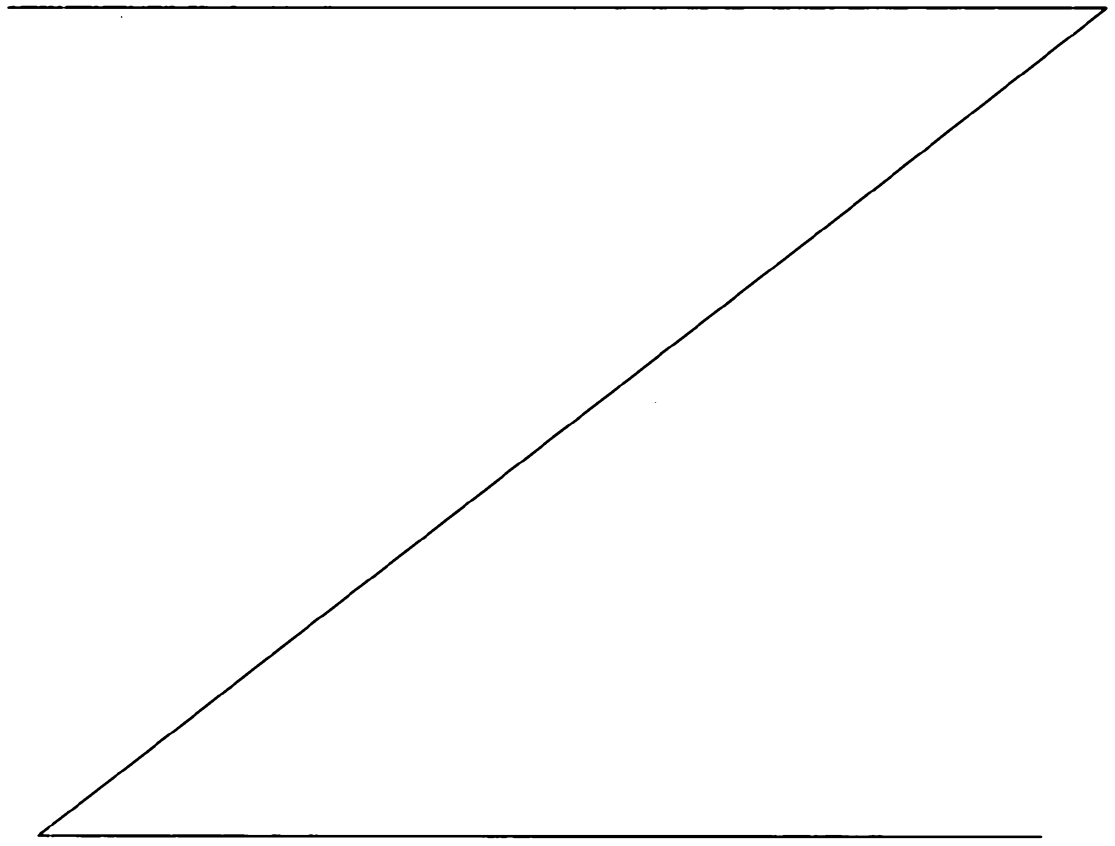
In a separate experiment organisms were grown on 2.5 g/L single sugars including cellobiose, glucose, xylose, galactose, arabinose, mannose and lactose as well as 5 g/L Avicel® and birchwood xylan. In Figure 18 product formation is compared on cellobiose and birchwood xylan after two days. Table 3 summarizes how bacterial strains may be categorized based on their substrate utilization.

Table 3

	cellobiose	glucose	xylose	galactose	arabinose	mannose	lactose
C.cellulolyticum	×	×	×	×	×		
C.stercorarium subs. leptospartum	×	×	×	×	×	×	×
C.kristjanssonii	×	×	×	×		×	×
C. phytofermentans	×	×	×	×		×	

Transgenic Conversion of Microorganisms

The present invention provides compositions and methods for the transgenic conversion of certain microorganisms. When genes encoding enzymes involved in the metabolic pathway of ethanol, including, for example, D-xylose and/or L-arabinose, are introduced into a bacterial strain that lacks one or more of these genes, for example, *C.*



thermocellum or *C. straminisolvens*, one may select transformed strains for growth on D-xylose or growth on L-arabinose. It is expected that genes from other Clostridial species should be expressed in *C. thermocellum* and *C. straminisolvens*. Target gene donors may include microorganisms that confer the ability to metabolize hexose and pentose sugars, *e.g.*, *C. cellulolyticum*, *Caldicellulosiruptor kristjanssonii*, *C. phytofermentans*, *C. stercorarium*, and *Thermoanaerobacterium saccharolyticum*.

The genomes of *T. saccharolyticum*, *C. cellulolyticum*, and *C. phytofermentans* are available. Accordingly, the present invention provides sequences which correspond to xylose isomerase and xylulokinase in each of the three hosts set forth above. In particular, the sequences corresponding to xylose isomerase (SEQ ID NO:6), xylulokinase (SEQ ID NO:7), L-arabinose isomerase (SEQ ID NO:8), and L-ribulose-5-phosphate 4-epimerase (SEQ ID NO:9) from *T. saccharolyticum* are set forth herein. Similarly, the sequences corresponding to xylose isomerase (SEQ ID NO:10), xylulokinase (SEQ ID NO:11), L-arabinose isomerase (SEQ ID NO:12), and L-ribulose-5-phosphate 4-epimerase (SEQ ID NO:13) from *C. cellulolyticum* are provided herein. *C. phytofermentans* utilizes the D-xylose pathway and does not utilize L-arabinose. Accordingly, the sequences corresponding to xylose isomerase (SEQ ID NO:14) and xylulokinase (SEQ ID NO:15) from *C. phytofermentans* are set forth herein.

C. kristjanssonii does metabolize xylose. To this end, the xylose isomerase (SEQ ID NO:71) and xylulokinase (SEQ ID NO:70) genes of *C. kristjanssonii* have been sequenced and are provided herein. *C. straminisolvens* has not been shown to grow on xylose, however it does contain xylose isomerase (SEQ ID NO:73) and xylulokinase (SEQ ID NO:72) genes, which may be functional after adaptation on xylose as a carbon source.

C. thermocellum and *C. straminisolvens* may lack one or more known genes or enzymes in the D-xylose to ethanol pathway and/or the L-arabinose utilization pathway. Figures 2 and 3 depict two key enzymes that are missing in each of these pathways in *C. thermocellum*. *C. straminisolvens* has xylose isomerase and xylulokinase, but the functionality of these enzymes is not known. Genomic sequencing has not revealed a copy of either L-arabinose isomerase or L-ribulose-5-phosphate 4-epimerase in *C. straminosolvens*.

C. thermocellum and *C. straminisolvens* are unable to metabolize xylulose which could reflect the absence (*C. thermocellum*) or lack of activity and/or expression (*C. straminsolvens*) of genes for xylose isomerase (referred to in Figure 2 as "XI" or 5.3.1.5), which converts D-xylose to D-xylulose, and xylulokinase (also referred to in Figure 2 as "XK" or 2.7.1.1), which converts D-xylulose to D-xylulose-5-phosphate. Furthermore, transport of xylose may be a limitation for *C. straminsolvens*. This potential limitation could be overcome by expression sugar transport genes from xylose utilizing organisms such as *T. saccharolyticum* and *C. kristjanssonii.*

C. thermocellum and *C. straminisolvens* are also unable to metabolize L-arabinose which could reflect the absence of genes for L-arabinose isomerase (also referred to in Figure 3 as 5.3.1.4) and L-ribulose-5-phosphate 4-epimerase (also referred to in Figure 3 as 5.1.3.4).

The four genes described above, *e.g.*, xylose isomerase, xylulokinase, L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase, are present in several Clostridial species and *Thermoanaerobacterium saccharolyticum* species, including, but not limited to, *Clostridium cellulolyticum* (see Figure 4), *Thermoanaerobacterium saccharolyticum*, *C. stercorarium*, *Caldiscellulosiruptor kristjanssonii*, and *C. phytofermentans*; these strains are good utilizers of these sugars. It will be appreciated that the foregoing bacterial strains may be used as donors of the genes described herein.

C. phytofermentans express the two xylose pathway genes described above (xylose isomerase and xylulokinase), but lack or do not express the arabinose pathway genes described above (L-arabinose isomerase and L-ribulose-5-phosphate 4-epimerase) (see Figure 5).

Accordingly, it is an object of the invention to modify some of the above-described bacterial strains so as to optimize sugar utilization capability by, for example, introducing genes for one or more enzymes required for the production of ethanol from biomass-derived pentoses, *e.g.*, D-xylose or L-arabinose metabolism. Promoters, including the native promoters of *C. thermocellum* or *C. straminisolvens*, such as triose phosphate isomerase (TPI), GAPDH, and LDH, may be used to express these genes. The sequences that correspond to native promoters of *C. thermocellum* include (TPI) (SEQ ID NO:16), GAPDH (SEQ ID NO:17), and LDH (SEQ ID NO:18). Once the gene has been cloned,

codon optimization may be performed before expression. Cassettes containing, for example, the native promoter, a xylanolytic gene or arabinolytic gene, and a selectable marker may then be used to transform *C. thermocellum* or *C. straminisolvens* and select for D-xylose and L-arabinose growth on medium containing D-xylose or L-arabinose as the sole carbohydrate source.

Transposons

To select for foreign DNA that has entered a host it is preferable that the DNA be stably maintained in the organism of interest. With regard to plasmids, there are two processes by which this can occur. One is through the use of replicative plasmids. These plasmids have origins of replication that are recognized by the host and allow the plasmids to replicate as stable, autonomous, extrachromosomal elements that are partitioned during cell division into daughter cells. The second process occurs through the integration of a plasmid onto the chromosome. This predominately happens by homologous recombination and results in the insertion of the entire plasmid, or parts of the plasmid, into the host chromosome. Thus, the plasmid and selectable marker(s) are replicated as an integral piece of the chromosome and segregated into daughter cells. Therefore, to ascertain if plasmid DNA is entering a cell during a transformation event through the use of selectable markers requires the use of a replicative plasmid or the ability to recombine the plasmid onto the chromosome. These qualifiers cannot always be met, especially when handling organisms that do not have a suite of genetic tools.

One way to avoid issues regarding plasmid-associated markers is through the use of transposons. A transposon is a mobile DNA element, defined by mosaic DNA sequences that are recognized by enzymatic machinery referred to as a transposase. The function of the transposase is to randomly insert the transposon DNA into host or target DNA. A selectable marker can be cloned onto a transposon by standard genetic engineering. The resulting DNA fragment can be coupled to the transposase machinery in an *in vitro* reaction and the complex can be introduced into target cells by electroporation. Stable insertion of the marker onto the chromosome requires only the function of the transposase machinery and alleviates the need for homologous recombination or replicative plasmids.

The random nature associated with the integration of transposons has the added advantage of acting as a form of mutagenesis. Libraries can be created that comprise

amalgamations of transposon mutants. These libraries can be used in screens or selections to produce mutants with desired phenotypes. For instance, a transposon library of a CBP organism could be screened for the ability to produce more ethanol, or less lactic acid and/or less acetate.

Native cellulolytic strategy

Naturally occurring cellulolytic microorganisms are starting points for CBP organism development via the native strategy. Anaerobes and facultative anaerobes are of particular interest. The primary objective is to engineer product yields and ethanol titers to satisfy the requirements of an industrial process. Metabolic engineering of mixed-acid fermentations in relation to these objectives has been successful in the case of mesophilic, non-cellulolytic, enteric bacteria. Recent developments in suitable gene-transfer techniques allow for this type of work to be undertaken with cellulolytic bacteria.

Recombinant cellulolytic strategy

Non-cellulolytic microorganisms with desired product-formation properties (*e.g.*, high ethanol yield and titer) are starting points for CBP organism development by the recombinant cellulolytic strategy. The primary objective of such developments is to engineer a heterologous cellulase system that enables growth and fermentation on pretreated lignocellulose. The heterologous production of cellulases has been pursued primarily with bacterial hosts producing ethanol at high yield (engineered strains of *E. coli*, *Klebsiella oxytoca*, and *Zymomonas mobilis*) and the yeast *Saccharomyces cerevisiae*. Cellulase expression in strains of *K. oxytoca* resulted in increased hydrolysis yields – but not growth without added cellulase – for microcrystalline cellulose, and anaerobic growth on amorphous cellulose. Although dozens of saccharolytic enzymes have been functionally expressed in *S. cerevisiae*, anaerobic growth on cellulose as the result of such expression has not been definitively demonstrated.

Aspects of the present invention relate to the use of thermophilic or mesophilic microorganisms as hosts for modification via the native cellulolytic strategy. Their potential in process applications in biotechnology stems from their ability to grow at relatively high temperatures with attendant high metabolic rates, production of physically and chemically stable enzymes, and elevated yields of end products. Major groups of thermophilic bacteria include eubacteria and archaebacteria. Thermophilic eubacteria

include: phototropic bacteria, such as cyanobacteria, purple bacteria, and green bacteria; Gram-positive bacteria, such as *Bacillus*, *Clostridium*, Lactic acid bacteria, and Actinomyces; and other eubacteria, such as *Thiobacillus*, Spirochete, *Desulfotomaculum*, Gram-negative aerobes, Gram-negative anaerobes, and *Thermotoga*. Within archaebacteria are considered Methanogens, extreme thermophiles (an art-recognized term), and *Thermoplasma*. In certain embodiments, the present invention relates to Gram-negative organotrophic thermophiles of the genera *Thermus*, Gram-positive eubacteria, such as genera *Clostridium*, and also which comprise both rods and cocci, genera in group of eubacteria, such as *Thermosipho* and *Thermotoga*, genera of Archaebacteria, such as *Thermococcus*, *Thermoproteus* (rod-shaped), *Thermofilum* (rod-shaped), *Pyrodictium, Acidianus, Sulfolobus, Pyrobaculum, Pyrococcus, Thermodiscus, Staphylothermus, Desulfurococcus, Archaeoglobus,* and *Methanopyrus*. Some examples of thermophilic or mesophilic (including bacteria, procaryotic microorganism, and fungi), which may be suitable for the present invention include, but are not limited to: *Clostridium thermosulfurogenes, Clostridium cellulolyticum, Clostridium thermocellum, Clostridium thermohydrosulfuricum, Clostridium thermoaceticum, Clostridium thermosaccharolyticum, Clostridium tartarivorum, Clostridium thermocellulaseum, Clostridium phytofermentans, Clostridium straminosolvens, Thermoanaerobacterium thermosaccarolyticum, Thermoanaerobacterium saccharolyticum, Thermobacteroides acetoethylicus, Thermoanaerobium brockii, Methanobacterium thermoautotrophicum, Anaerocellum thermophilium, Pyrodictium occultum, Thermoproteus neutrophilus, Thermofilum librum, Thermothrix thioparus, Desulfovibrio thermophilus, Thermoplasma acidophilum, Hydrogenomonas thermophilus, Thermomicrobium roseum, Thermus flavas, Thermus ruber, Pyrococcus furiosus, Thermus aquaticus, Thermus thermophilus, Chloroflexus aurantiacus, Thermococcus litoralis, Pyrodictium abyssi, Bacillus stearothermophilus, Cyanidium caldarium, Mastigocladus laminosus, Chlamydothrix calidissima, Chlamydothrix penicillata, Thiothrix carnea, Phormidium tenuissimum, Phormidium geysericola, Phormidium subterraneum, Phormidium bijahensi, Oscillatoria filiformis, Synechococcus lividus, Chloroflexus aurantiacus, Pyrodictium brockii, Thiobacillus thiooxidans, Sulfolobus acidocaldarius, Thiobacillus thermophilica, Bacillus stearothermophilus, Cercosulcifer hamathensis, Vahlkampfia reichi, Cyclidium citrullus, Dactylaria gallopava, Synechococcus lividus, Synechococcus elongatus, Synechococcus minervae, Synechocystis aquatilus, Aphanocapsa thermalis, Oscillatoria terebriformis,*

Oscillatoria amphibia, Oscillatoria germinata, Oscillatoria okenii, Phormidium laminosum, Phormidium parparasiens, Symploca thermalis, Bacillus acidocaldarias, Bacillus coagulans, Bacillus thermocatenalatus, Bacillus licheniformis, Bacillus pamilas, Bacillus macerans, Bacillus circulans, Bacillus laterosporus, Bacillus brevis, Bacillus subtilis, Bacillus sphaericus, Desulfotomaculum nigrificans, Streptococcus thermophilus, Lactobacillus thermophilus, Lactobacillus bulgaricus, Bifidobacterium thermophilum, Streptomyces fragmentosporus, Streptomyces thermonitrificans, Streptomyces thermovulgaris, Pseudonocardia thermophila, Thermoactinomyces vulgaris, Thermoactinomyces sacchari, Thermoactinomyces candidas, Thermomonospora curvata, Thermomonospora viridis, Thermomonospora citrina, Microbispora thermodiastatica, Microbispora aerata, Microbispora bispora, Actinobifida dichotomica, Actinobifida chromogena, Micropolyspora caesia, Micropolyspora faeni, Micropolyspora cectivugida, Micropolyspora cabrobrunea, Micropolyspora thermovirida, Micropolyspora viridinigra, Methanobacterium thermoautothropicum, Caldicellulosiruptor acetigenus, Caldicellulosiruptor saccharolyticus, Caldicellulosiruptor kristjanssonii, Caldicellulosiruptor owensensis, Caldicellulosiruptor lactoaceticus, variants thereof, and/or progeny thereof.

In certain embodiments, the present invention relates to thermophilic bacteria selected from the group consisting of *Fervidobacterium gondwanense, Clostridium thermolacticum, Moorella sp.*, and *Rhodothermus marinus*.

In certain embodiments, the present invention relates to thermophilic bacteria of the genera *Thermoanaerobacterium* or *Thermoanaerobacter*, including, but not limited to, species selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brockii*, variants thereof, and progeny thereof.

In certain embodiments, the present invention relates to microorganisms of the genera *Geobacillus, Saccharococcus, Paenibacillus, Bacillus*, and *Anoxybacillus*, including, but not limited to, species selected from the group consisting of: *Geobacillus*

thermoglucosidasius, Geobacillus stearothermophilus, Saccharococcus caldoxylosilyticus, Saccharoccus thermophilus, Paenibacillus campinasensis, Bacillus flavothermus, Anoxybacillus kamchatkensis, Anoxybacillus gonensis, variants thereof, and progeny thereof.

In certain embodiments, the present invention relates to mesophilic bacteria selected from the group consisting of *Saccharophagus degradans; Flavobacterium johnsoniae; Fibrobacter succinogenes; Clostridium hungatei*; *Clostridium phytofermentans*; *Clostridium cellulolyticum*; *Clostridium aldrichii*; *Clostridium termitididis*; *Acetivibrio cellulolyticus*; *Acetivibrio ethanolgignens*; *Acetivibrio multivorans*; *Bacteroides cellulosolvens*; and *Alkalibacter saccharofomentans*, variants thereof and progeny thereof.

Methods of the Invention

During glycolysis, cells convert simple sugars, such as glucose, into pyruvic acid, with a net production of ATP and NADH. In the absence of a functioning electron transport system for oxidative phosphorylation, at least 95% of the pyruvic acid is consumed in short pathways which regenerate NAD^+, an obligate requirement for continued glycolysis and ATP production. The waste products of these NAD^+ regeneration systems are commonly referred to as fermentation products.

Microorganisms produce a diverse array of fermentation products, including organic acids, such as lactate (the salt form of lactic acid), acetate (the salt form of acetic acid), succinate, and butyrate, and neutral products, such as ethanol, butanol, acetone, and butanediol. End products of fermentation share to varying degrees several fundamental features, including: they are relatively nontoxic under the conditions in which they are initially produced, but become more toxic upon accumulation; and they are more reduced than pyruvate because their immediate precursors have served as terminal electron acceptors during glycolysis. Aspects of the present invention relate to the use of gene knockout technology to provide novel microorganisms useful in the production of ethanol from lignocellulosic biomass substrates. The transformed organisms are prepared by deleting or inactivating one or more genes that encode competing pathways, such as the non-limiting pathways to organic acids described herein, optionally followed by a growth-based selection for mutants with improved performance for producing ethanol as a fermentation product.

In certain embodiments, a thermophilic or mesophilic microorganism, which in a native state contains at least one gene that confers upon the microorganism an ability to produce lactic acid as a fermentation product, is transformed to decrease or eliminate expression of said at least one gene. The gene that confers upon said microorganism an ability to produce lactic acid as a fermentation product may code for expression of lactate dehydrogenase. The deletion or suppression of the gene(s) or particular polynucleotide sequence(s) that encode for expression of LDH diminishes or eliminates the reaction scheme in the overall glycolytic pathway whereby pyruvate is converted to lactic acid; the resulting relative abundance of pyruvate from these first stages of glycolysis should allow for the increased production of ethanol.

In certain embodiments, a thermophilic or mesophilic microorganism, which in a native state contains at least one gene that confers upon the microorganism an ability to produce acetic acid as a fermentation product, is transformed to eliminate expression of said at least one gene. The gene that confers upon the microorganism an ability to produce acetic acid as a fermentation product may code for expression of acetate kinase and/or phosphotransacetylase. The deletion or suppression of the gene(s) or particular polynucleotide sequence(s) that encode for expression of ACK and/or PTA diminishes or eliminates the reaction scheme in the overall glycolytic pathway whereby acetyl CoA is converted to acetic acid (Figure 1); the resulting relative abundance of acetyl CoA from these later stages of glycolysis should allow for the increased production of ethanol.

In certain embodiments, the above-detailed gene knockout schemes can be applied individually or in concert. Eliminating the mechanism for the production of lactate (*i.e.*, knocking out the genes or particular polynucleotide sequences that encode for expression of LDH) generates more acetyl CoA; it follows that if the mechanism for the production of acetate is also eliminated (*i.e.*, knocking out the genes or particular polynucleotide sequences that encode for expression of ACK and/or PTA), the abundance of acetyl CoA will be further enhanced, which should result in increased production of ethanol.

In certain embodiments, it is not required that the thermophilic or mesophilic microorganisms have native or endogenous PDC or ADH. In certain embodiments, the genes encoding for PDC and/or ADH can be expressed recombinantly in the genetically modified microorganisms of the present invention. In certain embodiments, the gene knockout technology of the present invention can be applied to recombinant

microorganisms, which may comprise a heterologous gene that codes for PDC and/or ADH, wherein said heterologous gene is expressed at sufficient levels to increase the ability of said recombinant microorganism (which may be thermophilic) to produce ethanol as a fermentation product or to confer upon said recombinant microorganism (which may be thermophilic) the ability to produce ethanol as a fermentation product.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 70% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 80% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, aspects of the present invention relate to fermentation of lignocellulosic substrates to produce ethanol in a concentration that is at least 90% of a theoretical yield based on cellulose content or hemicellulose content or both.

In certain embodiments, substantial or complete elimination of organic acid production from microorganisms in a native state may be achieved using one or more site-directed DNA homologous recombination events.

Operating either a simultaneous saccharification and co-fermentation (SSCF) or CBP process at thermophilic temperatures offers several important benefits over conventional mesophilic fermentation temperatures of 30-37 °C. In particular, costs for a process step dedicated to cellulase production are substantially reduced (*e.g.*, 2-fold or more) for thermophilic SSCF and are eliminated for CBP. Costs associated with fermentor cooling and also heat exchange before and after fermentation are also expected to be reduced for both thermophilic SSCF and CBP. Finally, processes featuring thermophilic biocatalysts may be less susceptible to microbial contamination as compared to processes featuring conventional mesophilic biocatalysts.

The ability to redirect electron flow by virtue of modifications to carbon flow has broad implications. For example, this approach could be used to produce high ethanol yields in strains other than *T. saccharolyticum* and/or to produce solvents other than ethanol, for example, higher alcohols (i.e., butanol).

Metabolic engineering through antisense oligonucleotide (asRNA) strategies

Fermentative microorganisms such as yeast and anaerobic bacteria ferment sugars to ethanol and other reduced organic end products. Theoretically, carbon flow can be directed to ethanol production if the formation of competing end-products, such as lactate and acetate, can be suppressed. The present invention provides several genetic engineering approaches designed to remove such competing pathways in the CBP organisms of the invention. The bulk of these approaches utilize knock-out constructs (for single crossover recombination) or allele-exchange constructs (for double crossover recombination) and target the genetic loci for *ack* and *ldh*. Although these tools employ "tried and true" strain development techniques, there are several potential issues that could stall progress: (i) they are dependent on the host recombination efficiency which in all cases is unknown for the CBP organisms; (ii) they can be used to knock out only one pathway at a time, so successive genetic alterations are incumbent upon having several selectable markers or a recyclable marker; (iii) deletion of target genes may be toxic or have polar effects on downstream gene expression.

The present invention provides additional approaches towards genetic engineering that do not rely on host recombination efficiency. One of these alternative tools is called antisense RNA (asRNA). Although antisense oligonucleotides have been used for over twenty-five years to inhibit gene expression levels both *in vitro* and *in vivo*, recent advances in mRNA structure prediction has facilitated smarter design of asRNA molecules. These advances have prompted a number of groups to demonstrate the usefulness of asRNA in metabolic engineering of bacteria.

The benefits of using asRNA over knock-out and allele-exchange technology are numerous: (i) alleviates the need for multiple selectable markers because multiple pathways can be targeted by a single asRNA construct; (ii) attenuation level of target mRNA can be adjusted by increasing or decreasing the association rate between asRNA; (iii) pathway inactivation can be conditional if asRNA transcripts are driven by conditional promoters. Recently, this technology has been used to increase solventogenesis in the Gram positive mesophile, *Clostridium acetobutylicum* (Tummala et al. (2003)). Although the exact molecular mechanism of how asRNA attenuates gene expression is unclear, the likely mechanism is triggered upon hybridization of the asRNA to the target mRNA. Mechanisms may include one or more of the following: (i) inhibition of translation of

mRNA into protein by blocking the ribosome binding site from properly interacting with the ribosome, (ii) decreasing the half-life of mRNA through dsRNA-dependent RNases, such as RNase H, that rapidly degrade duplex RNA, and (iii) inhibition of transcription due to early transcription termination of mRNA.

Design of antisense sequences

asRNAs are typically 18-25 nucleotides in length. There are several computation tools available for rational design of RNA-targeting nucleic acids (Sfold, Integrated DNA Technologies, STZ Nucleic Acid Design) which may be used to select asRNA sequences. For instance, the gene sequence for *Clostridium thermocellum ack* (acetate kinase) can be submitted to a rational design server and several asRNA sequences can be culled. In brief, the design parameters select for mRNA target sequences that do not contain predicted secondary structure.

Design of delivery vector

A replicative plasmid will be used to deliver the asRNA coding sequence to the target organism. Vectors such as, but not limited to, pNW33N, pJIR418, pJIR751, and pCTC1, will form the backbone of the asRNA constructs for delivery of the asRNA coding sequences to inside the host cell. In addition to extra-chromosomal (plasmid based) expression, asRNAs may be stably inserted at a heterologous locus into the genome of the microorganism to get stable expression of asRNAs. In certain embodiments, strains of thermophilic or mesophilic microorganisms of interest may be engineered by site directed homologous recombination to knockout the production of organic acids and other genes of interest may be partially, substantially, or completely deleted, silenced, inactivated, or down-regulated by asRNA.

Promoter choice

To ensure expression of asRNA transcripts, compatible promoters for the given host will be fused to the asRNA coding sequence. The promoter-asRNA cassettes are constructed in a single PCR step. Sense and antisense primers designed to amplify a promoter region will be modified such that the asRNA sequence (culled from the rational design approach) is attached to the 5' end of the antisense primer. Additionally, restriction sites, such as *Eco*RI or *Bam*HI, will be added to the terminal ends of each primer so that the

final PCR amplicon can be digested directly with restriction enzymes and inserted into the vector backbone through traditional cloning techniques.

With respect to microorganisms that do not have the ability to metabolize pentose sugars, but are able to metabolize hexose sugars as described herein, it will be appreciated that the *ack* and *ldh* genes of *Clostridium thermocellum* and *Clostridium straminisolvens*, for example, may be targeted for inactivation using antisense RNA according to the methods described herein.

With respect to microorganisms that confer the ability to metabolize pentose and hexose sugars as described herein, it will be appreciated that the *ack* and *ldh* genes of *Clostridium cellulolyticum*, *Clostridium phytofermentans* and *Caldicellulosiruptor kristjanssonii*, for example, may be targeted for inactivation using antisense according to the methods described herein.

In addition to antibiotic selection for strains expressing the asRNA delivery vectors, such strains may be selected on conditional media that contains any of the several toxic metabolite analogues such as sodium fluoroacetate (SFA), bromoacetic acid (BAA), chloroacetic acid (CAA), 5-fluoroorotic acid (5-FOA) and chlorolactic acid. Use of chemical mutagens including, but not exclusively, ethane methyl sulfonate (EMS) may be used in combination with the expression of antisense oligonucleotide (asRNA) to generate strains that have one or more genes partially, substantially, or completely deleted, silenced, inactivated, or down-regulated.

EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

EXAMPLE 1

Generation of Custom Transposons For Mesophilic and Thermophilic Cellulolytic, Xylanolytic Organisms

The present invention provides methods for generating custom transposons for cellulolytic and/or xylanolytic and/or thermophilic organisms. To do this, a native promoter from the host organism will be fused to a selectable marker which has been determined to work in this organism. This fragment will be cloned into the EZ-Tn5™ transposon that is carried on the vector pMOD™-2<MCS> (Epicenter®Biotechnologies). For example, the *C. thermocellum* the *gapDH* promoter will be fused to the mLs drug marker, as well as the *cat* gene and then subcloned into vector pMOD™-2<MCS>.

Commercial transposons are lacking in thermostable drug markers and native promoters of cellulolytic and/or xylanolytic and/or thermophilic organisms. The mLs and *cat* markers have functioned in thermophilic bacteria and the *gapDH* promoter regulates a key glycolytic enzyme and should be constantly expressed. The combination of the above drug markers and the *gapDH* promoter will greatly enhance the probability of generating a functional transposon. This approach may be applied to other cellulolytic and/or xylanolytic and/or thermophilic organisms.

Experimental Design

Figure 26 is a diagram taken from the Epicenter®Biotechnologies user manual, which is incorporated herein by reference, representing bp 250-550 of pMOD™-2<MCS>. In the top portion, the black arrowheads labeled ME denote 19 bp mosaic ends that define the transposon. The *Eco*RI and *Hin*dIII sites define the multi-cloning site, which is represented by the black box labeled MCS. In the bottom portion, the DNA sequence and the restriction enzymes associated with the MCS are shown.

The following primers will be used to amplify promoter fusion fragments from pMQ87-*gapDH-cat* and pMQ87-*gapDH*-mls: GGCGgaattc CTT GGT CTG ACA ATC GAT GC (SEQ ID NO:19); GGCGgaattc TATCAGTTATTACCCACTTTTCG (SEQ ID NO:20). The lower case letters denote engineered *Eco*RI restriction sites. The size of the amplicon generated will be ~1.9 kb. Standard molecular procedures will allow the amplicon to be digested with *Eco*RI and cloned into the unique *Eco*RI site of pMOD™-

2<MCS>. The transposon and subsequent transpososome will be generated and introduced into host organisms as described by the manufacturer.

EXAMPLE 2

Constructs for Engineering Cellulolytic and Xylanolytic Strains

The present invention provides compositions and methods for genetically engineering an organism of interest to CBP by mutating genes encoding key enzymes of metabolic pathways which divert carbon flow away from ethanol. Single crossover knockout constructs are designed so as to insert large fragments of foreign DNA into the gene of interest to partially, substantially, or completely delete, silence, inactivate, or down-regulate it. Double crossover knockout constructs are designed so as to partially, substantially, or completely delete, silence, inactivate, or down-regulate the gene of interest from the chromosome or replace the gene of interest on the chromosome with a mutated copy of the gene, such as a form of the gene interrupted by an antibiotic resistance cassette.

The design of single crossover knockout vectors requires the cloning of an internal fragment of the gene of interest into a plasmid based system. Ideally, this vector will carry a selectable marker that is expressed in the host strain but will not replicate in the host strain. Thus, upon introduction into the host strain the plasmid will not replicate. If the cells are placed in a conditional medium that selects for the marker carried on the plasmid, only those cells that have found a way to maintain the plasmid will grow. Because the plasmid is unable to replicate as an autonomous DNA element, the most likely way that the plasmid will be maintained is through recombination onto the host chromosome. The most likely place for the recombination to occur is at a region of homology between the plasmid and the host chromosome.

Alternatively, replicating plasmids can be used to create single crossover interruptions. Cells that have taken up the knockout vector can be selected on a conditional medium, then passaged in the absence of selection. Without the positive selection provided by the conditional medium, many organisms will lose the plasmid. In the event that the plasmid is inserted onto the host chromosome, it will not be lost in the absence of selection. The cells can then be returned to a conditional medium and only those that have retained the marker, through chromosomal integration, will grow. A PCR based method will be devised to screen for organisms that contain the marker located on the chromosome.

The design of double crossover knockout vectors requires at least cloning the DNA flanking (~ 1 kb) the gene of interest into a plasmid and in some cases may include cloning the gene of interest. A selectable marker may be placed between the flanking DNA or if the gene of interest is cloned the marker is placed internally with respect to the gene. Ideally the plasmid used is not capable of replicating in the host strain. Upon the introduction of the plasmid into the host and selection on a medium conditional to the marker, only cells that have recombined the homologous DNA onto the chromosome will grow. Two recombination events are needed to replace the gene of interest with the selectable marker.

Alternatively, replicating plasmids can be used to create double crossover gene replacements. Cells that have taken up the knockout vector can be selected on a conditional medium, then passaged in the absence of selection. Without the positive selection provided by the conditional medium, many organisms will lose the plasmid. In the event that the drug marker is inserted onto the host chromosome, it will not be lost in the absence of selection. The cells can then be returned to a conditional medium and only those that have retained the marker, through chromosomal integration, will grow. A PCR based method may be devised to screen for organisms that contain the marker located on the chromosome.

In addition to antibiotic selection schemes, several toxic metabolite analogues such as sodium fluoroacetate (SFA), bromoacetic acid (BAA), chloroacetic acid (CAA), 5-fluoroorotic acid (5-FOA) and chlorolactic acid may be used to select mutants arising from either homologous recombinations, or transposon-based strategies. Use of chemical mutagens including, but not exclusively, ethane methyl sulfonate (EMS) may be used in combination with the directed mutagenesis schemes that employ homologous recombinations, or transposon-based strategies.

C. cellulolyticum Knockout Constructs

Acetate kinase (gene 131 from C. cellulolyticum published genome):

Single Crossover

The acetate kinase gene of *C. cellulolyticum* is 1,110 bp in length. A 662 bp internal fragment (SEQ ID NO:21) spanning nucleotides 91-752 was amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers. Selectable markers may include those that provide erythromycin and chloramphenicol

resistance. These plasmids will be used to disrupt the *ack* gene. A map of the *ack* gene and the region amplified by PCR for gene disruption are shown in Figure 19. The underlined portions of SEQ ID NO:21 set forth below correspond to the sites that are *Eco*RI sites that flank the knockout fragment.

gaattctgcgacagaatagggattgacaattcctttataaagcaatcaaggggttcagaagaggctgttattttgaataaagagctaaagaat
cacaaagatgcaatagaggctgttatttctgcactgactgacgataatatgggcgttataaaaaacatgtccgaaatatcagcagtgggaca
cagaatagtacacggcggtgaaaaattcaacagttctgtagttatagatgaaaacgttatgaatgcagtaagagagtgtatagacgttgcac
cgcttcataatccgccgaatattataggtatagaggcttgccagcagattatgcccaatatacctatggtagctgtatttgataccactttccaca
gctccatgcctgattatgcatacctttacgcattgccatatgaactttatgaaaagtacggtataagaaaatatggtttccacggaacatcacac
aaatatgttgcagaaagagcttctgcaatgcttgataagtctttgaacgaattaaagataattacatgccatcttgggaacggttcaagtatttgt
gctgttaacaagggtaaatcaattgatacttccatgggctttacacctttgcagggacttgcaatgggtacaagaagcggtacaatagaccct
gaagttgttacgaattc

These sites were engineered during the design of the "ack KO primers" and will allow subsequent cloning of the fragment into numerous vectors.

Double crossover

To construct a double crossover vector for the *ack* gene of *C. cellulolyticum* ~1 kb of DNA flanking each side of the *ack* gene will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers may include those that provide erythromycin and chloramphenicol resistance. The 3' flanking region of the *ack* gene is not available in the available draft genome. To acquire this DNA, a kit such as GenomeWalker from Clontech will be used.

Lactate dehydrogenase (genes 2262 and 2744 of C. cellulolyticum published genome):

Single crossover

The *ldh* genes of *C. cellulolyticum* are 951 bp (for gene 2262) (SEQ ID NO:22) and 932 bp (for gene 2744) (SEQ ID NO:23) in length. A ~500 bp internal fragment near the 5' end of each gene will be amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers. Selectable markers may include those that provide drug resistance, such as erythromycin and chloramphenicol. These plasmids will be used to disrupt the *ldh* 2262 and *ldh* 2744 genes. As an example, a map of the *ldh* 2262 gene and the region amplified by PCR for gene disruption are shown in Figure 20.

Double Crossover

To construct a double crossover knockout vector for the *ldh* gene(s) of *C. cellulolyticum* ~1 kb of DNA flanking each side of the *ldh* gene(s) will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers may include those that provide drug resistance, such as erythromycin and chloramphenicol. Figure 21 provides an example of *C. cellulolyticum ldh* (2262) double crossover knock out fragment.

In the sequence set forth below (SEQ ID NO:24) the mL*s* gene (selectable marker) is underlined and the flanking DNA is the remaining sequence. During primer design, restriction sites will be engineered and the 5' and 3' ends of the above fragment so that it can be cloned into a number of replicative and non-replicative vectors. The same strategy will be used to create a vector to delete *ldh* 2744.

gacgcatacaggttgtaacacccatttcccttagcttttcgggagatgaataaaacaaactttccgggtcctttaccacaccgcccacataaa
gagctatgccgcatgaaagaaacgatatgttatcatttttttcgtaaactgttatttccgaacccggataaagctttaccatattattaactgctgcc
gtccctgcatgtgtacaccctataaccactattttcatatacatcctcctttgtttgcttgtaaatatatcccatatataccacctaaatatattttataa
acaaattcggtatatcattcttttggtaaataaaaagtacatccgatattagaatgtacctaaaaaaaaattattattttattgtatatgctttatctgtttt
cattatatggtttgctatccattctacggtaaaatcaagtaattccattaagtactgatcctgatccttgtctatcctgctataatccgtattactgatttt
ctcaataaaatcatggtgttcaactttgtgggagagaagcttgcgatatcctatgctatgcatgtattcttcttcataggtaaaatgaaagacagt
gtaatcttttagttccgtaattagccgtacaatttcatcatatttgtctgtaataagctgatttttcgtggcctcataaatttccgaagcaatctggaat
agtttcttatgctgttcgtcgattttctcaattccaagaataaattcgtctctccattctatcatatggaccctcctaaattgtaatgtataccaagatta
tacatacttcctagaatataaacaatacaaggataaaattttaatatcgtatacctacataaatgactaacttaaagctctctaaaacttctttttta
ttatttctatactactaaaatcaaaaatattctctaaagtatttctacaaatgttgtttttgcaacaaagtagtatacttttgcacccagaatgttttgtta
taacttacaaattaggggtatatttatagtaaatactaaatggaagagtaggatattgatt<u>atgaacgagaaaaatataaaacacagtcaaa</u>
<u>actttattacttcaaaacataatatagataaaataatgacaaatataagattaaatgaacatgataatatctttgaaatcggctcaggaaaagg</u>
<u>gcattttacccttgaattagtacagaggtgtaatttcgtaactgccattgaaatagaccataaattatgcaaaactacagaaaataaacttgttg</u>
<u>atcacgataatttccaagttttaaacaaggatatattgcagtttaaatttcctaaaaaccaatcctataaaatatttggtaatatacctataacata</u>
<u>agtacggatataatacgcaaaattgtttttgatagtatagctgatgagatttatttaatcgtggaatacgggtttgctaaaagattattaaatacaa</u>
<u>aacgctcattggcattatttttaatggcagaagttgatatttctatattaagtatggttccaagagaatattttcatcctaaacctaaagtgaatagct</u>
<u>cacttatcagattaaatagaaaaaaatcaagaatatcacacaaagataaacagaagtataattatttcgttatgaaatgggttaacaaagaa</u>
<u>tacaagaaaatatttacaaaaaatcaatttaacaattccttaaaacatgcaggaattgacgatttaaacaatattagctttgaacaattcttatct</u>
<u>cttttcaatagctataaattatttaataag</u>atccccttacttcggatgcatgccgcaggcaggcatccgaagtagtttctccattatacaagtattc
tcttgagtacgtcgtcgcttctcagcagctgctttgctttttccctgttttccggcacatggagataagtgtatctgttaggcttaatagtgtgtgccat
gtcaattgccttttcgaagtcatctgccttcatttttaaggtttccacaaaattgataaaacccgtatcagtcagaaattttactacccgctgatatct
gtgttcttgaaccctgctcataagatagg ttgcaatcccaacctgaattccatgaagctgaggtgtctccagcagcttatctaaagcatgagat
attagatgctcactaccgctggctggagcactgctgtctgctatctgcatggcaattccgctcattgtcagagagtctaccatttccttaaaaag
aagttttctgtaacctgtgtgtagggcatccttacaatactgtttactgacttttagcaatcattgcagcaaaatcgtcaacctttgccgcattgttc
ctttcttcaaaataccagtcatacacagccgtaattttggatattatgtctccgagacctgaataaataaatttcataggtgcatttttttaatacatct
aaatccactaatattccaaatggcatcgaggcatgtacggaagtacgcctgccatttataatcaaagagcagcctgagctggaaaaaccat
cgtttgaggttgatgtaggtatactgataaaaggaagcttgtttaaaaaagctatatatttggctgcatcaagcacctttcctcctcctactccgac
cactgcatcggttttggagggaatagtaaaagccttgagcataagattttcaagctttatgtcatcatagtcgtaagtttcaagtactgcaagag
attttcttgactttatggaatccagaatcttttcaccaaataagtcacgtattccctctccaaaaagtactacaacattactaattcctgccctttcaa
tatgtgc

C. phytofermentans Knockout Constructs

For acetate kinase (gene 327 from C. phytofermentans published genome):

Single crossover

The acetate kinase gene of *C. phytofermentans* is 1,244 bp in length. A 572 bp internal fragment spanning nucleotides 55-626 will be amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers. Selectable markers to use will include those that provide drug resistance to *C. phytofermentans*. These plasmids will be used to disrupt the *ack* gene. A map of the *ack* gene and the region amplified by PCR for gene disruption are shown in Figure 22. Restriction sites will be engineered during the design of the "ack KO primers" and will allow subsequent cloning of the fragment into numerous vectors. The sequence of the knockout fragment described above is set forth as SEQ ID NO:25.

Double crossover

To construct a double crossover knockout vector for the *ack* gene of *C. phytofermentans* ~1 kb of DNA flanking each side of the *ack* gene will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers to use will include those that provide drug resistance to this strain. An example of a putative double crossover knockout construct with the mLs gene as a putative selectable marker is shown in Figure 23.

The sequence that corresponds to the fragment depicted in Figure 23 (SEQ ID NO:26) is set forth below. The mL*s* gene (putative selectable marker) is underlined and the remainder of the sequence corresponds to the flanking DNA. During primer design, restriction sites will be engineered and the 5' and 3' ends of the above fragment so that it can be cloned into a number of replicative and non-replicative vectors.

ctgagtgcaatgtaaaaaaggatgcctcaagtattcttgaaacatccttatattatactacaaaatcataaagtaaattactcagctgtagcaat
gatctctttttgttgtaagatccacaagctttacaaactctatgaggcatcataagtgcaccacacttgctgcatttcactaagtttggagcagtc
atcttccagtttgcacgacgactatctcttctagctttggaatgtttattctttggacaaatagctcccattgattacacctccttaaacttgttaaaaa
tatctcggatagcagacattcttgggtctagttctgtacggtcacacccgcactctccttcatttaggttagcaccgcagaccttgcagattccttt
acagtcttctttgcacagaaccttcattgggaaaccaatcaagacttcttcatagataagtttatctacgtctaaatcatatccggaaacaaaatt
tgtttcatctaaatcctcggtacgctgttcctctgttttcgatacatcaatctctgtagccacgtcgatgtcttgttggatggtttcttccttcaaacaac
gatcgcaaggaacggctaacgctaatttcgtttttgcttccaccagaatttttcggccacctagattagttaatctaagtttaaccggttctttatag
gtaatagaataaccgacaccatttaattcgaatatatcaaattcaatcggtgcagtgtattctttgagaccattaggaacattcatgacttcaga
catttgtatcagcataagtaactcctgtctaaaaaaacgcataatgtaagcgcccaaaaattcacactgttagtattataaacgcttaaaatag

gtttgtcaactcctaactgttaaaaatgtcagaattgtgtaaccatattttctcttcattatcgttcttcccttattaaataatttatagctattgaaaaga
gataagaattgttcaaagctaatattgtttaaatcgtcaattcctgcatgttttaaggaattgttaaattgatttttttgtaaatattttcttgtattctttgtta
acccatttcataacgaaataattatacttctgtttatctttgtgtgatattcttgatttttttctatttaatctgataagtgagctattcactttaggtttaggat
gaaaatattctcttggaaccatacttaatatagaaatatcaacttctgccattaaaaataatgccaatgagcgttttgtatttaataatcttttagca
aacccgtattccacgattaaataaatctcatcagctatactatcaaaaacaattttgcgtattatatccgtacttatgttataaggtatattaccaaa
tattttataggattggtttttaggaaatttaaactgcaatatatccttgtttaaaacttggaaattatcgtgatcaacaagtttattttctgtagttttgcat
aatttatggtctatttcaatggcagttacgaaattacacctctgtactaattcaagggtaaaatgccctttcctgagccgatttcaaagatattatc
atgttcatttaatcttatatttgtcattattttatctatattatgttttgaagtaataaagttttgactgtgttttatatttttctcgttcattgtatttctccttataat
gttcttaaattcatttatcacggggcaacttaatatatccgaaatatagttcttctatatcgttcccccagtataatgattattatactatttaatcttcaa
cttaacaattggagtttccagttaagaaataataatttaatgccaaagcggatattcgcaatccgcttacgctacttgctcataacctcaacagg
caatgaagctaagttaattatttactctgtgcctgaacagcagtgattgcaacaacaccaacgatatcatcagaagaacaacctcttgataaa
tcatttactggagctgcaataccctgagttaatggtccataagcttctgcctttgcaagacgctgtgttaacttatatccaatgttaccagcatcaa
ggtctgggaagattaatacgttagctttccagcaatatcactaccaggagcttttgaagcacctacactaggaacgattgctgcatctaactg
gaactcgccgtcgatcttatattctgggtataattcatttgcaatcttagttgcttctacaaccttatcaacatctgcatgctttgcgcttccctttgttga
atgagaaagcatagctacgataggttcagagccaactaattgttcaaaactcttcgctgtggaaccagcgattgctgctaactcttcagcattt
ggattctgatttaaaccagcatcagagaaaaggaaagttccatttgcgcccatatcacaattaggtactaccattacgaagaaagcagaaa
ctaacttagtatttggagcagttttaaaatctgaagacatggtcttaaggtatctgctgtagagtgacaagcaccagatactaaaccatctgca
tcgcccatcttaaccatcattacaccgtatgtaatgtagtctgttgttaaaagctcttttgctttttcaggggtcatgcctttgcctgtctaagttctaca
agcttgttaatgtaagc

For Lactate dehydrogenase (genes 1389 and 2971 of C. phytofermentans published genome)

Single crossover

The *ldh* genes of *C. phytofermentans* are 978 bp (for gene 1389) (SEQ ID NO:27) and 960 bp (for gene 2971) (SEQ ID NO:28) in length. A ~500 bp internal fragment near the 5' end of each gene will be amplified by PCR and cloned into suicide vectors and replicating vectors that have different selectable markers. Selectable markers to use will include those that provide drug resistance. These plasmids will be used to disrupt the *ldh* 1389 and *ldh* 2971 genes. As an example, a map of the *ldh* 1389 gene and the region amplified by PCR for gene disruption are shown in Figure 24.

Double crossover

To construct a double crossover knockout vector for the *ldh* gene(s) of *C. phytofermentans* ~1 kb of DNA flanking each side of the *ldh* gene(s) will be cloned. A selectable marker will be inserted between the flanking DNA. Selectable markers to use will include those that provide drug resistance to this strain. An example of a putative double crossover knockout construct with the mLs gene as a putative selectable marker is shown in Figure 25.

The sequence that corresponds to the fragment depicted in Figure 25 is set forth below as SEQ ID NO:29. The mL*s* gene (selectable marker) is underlined and the remaining portion of the sequence corresponds to the flanking DNA. During primer design, restriction sites will be engineered and the 5' and 3'ends of the above fragment so that it can be cloned into a number of replicative and non-replicative vectors. The same strategy will be used to create a vector to delete *ldh* 2971.

tggaatctcactatgcaccaatgtggtactaaattatatctttatctatggaaaattaggttttccgcgaatggagatagagggagctgccattgc
tactttaatttgtagaattcttgagagtattttagttgttatttatatgtataagggtgagaaggtacttaagatgagactttcttatatttttaagagatct
aaacagtattttcgctctttggctcgttatagtgcgccagtgcttatgagtgaggttaactgggggcttgggattgctgttcagtctgcaatcattgg
gcgtatgggtgttagttttcttacagccgccagcttcattaatgtagtacaacagttagccggaatcattctgattggtattggtgtgggttcgagc
attataatagggaatttgattggtgagggaaaagagcatgaggcgagaatgctagccaataagttaatacgtatcagtatgatactcggagg
aattgttgcttttgcagtaatcttactacgtccaatcgctcctaactttattgaggcgtctaaggaaacagcggatttaattcgtcagatgctatttgtt
tcggcttacctcttattcttccaagccttatctgtattaactatggccggaatattacgtggtgcaggggataccctttactgtgcaacctttgatgttt
tgaccttatgggtactaaaacttggaggaggtttgcttgcaaccatagtacttcatcttccacctgtatgggtttactttatcttaagtagcgatgag
tgtgttaaagcgctatttacggtaccgcgggtcttaaagggacgttggattcatgatacaacactgcattaagatttcatatgtccagatattttg
cacagtagcataattactagagcttattcctataatattcataggttttgatggtccattttacgttacgatagcatatattacatcaaaaccaattct
atataagatgaggttatagtatgaacgagaaaaatataaaacacagtcaaaactttattacttcaaaacataatatagataaaataatgaca
aatataagattaaatgaacatgataatatctttgaaatcggctcaggaaaagggcattttaccttgaattagtacagaggtgtaatttcgtaac
tgccattgaaatagaccataaattatgcaaaactacagaaaataaacttgttgatcacgataatttccaagttttaaacaaggatatattgcagt
ttaaatttcctaaaaaccaatcctataaaatatttggtaatatacctataacataagtacggatataatacgcaaaattgtttttgatagtatagct
gatgagatttatttaatcgtggaatacgggtttgctaaaagattattaaatacaaaacgctcattggcattatttttaatggcagaagttgatatttct
atattaagtatggttccaagagaatattttcatcctaaacctaaagtgaatagctcacttatcagattaaatagaaaaaaatcaagaatatcac
acaaagataaacagaagtataattatttcgttatgaaatgggttaacaaagaatacaagaaaatatttacaaaaaatcaatttaacaattcctt
aaaacatgcaggaattgacgatttaaacaatattagctttgaacaattcttatctcttttcaatagctataaattatttaataagaagtaataggaa
ataatactcgaattattctgcaatctgttctaaaaaataaaattaagaaattactatagcaagccaggttaaaattactagcttgctatttttgtgca
tttagtacagttttgattattaaagaataaatttaataactattttgcaataagttattgactatttcacaagttagtgttactatacaagtatgaaata
aagatacataaaaaaataaataatatgaaacataaattcatgacatgcggaatagaatgaaagaatattatgtcggttcctaatactaaatg
gatataacaatctattgaaacacttatggggtgtaagtgtggagagaatttctaaagcgccaaaagactctacatatgaaattctaaagcttca
cacgggaataatctaatttatgtatcttattatcataattcaggaaggtagtgtgaaaatataaaaattagttttcctgtttcattcaggcagtagca
tttcttaaacaaatttgctatgcattgggtgttatctgaaaaacaaaaagcaattttctcacaacttatttctgaacaacaatggtattaaaaatttg
gaggaggattttactatgaaaaaaacggtaacattactgttggttctgaccatggtggtaagcttatttgcagcatgtggtaagaaaaatggat
caagcgaaaccggcacaaaagatcctgtggcaacaagcggtgcaaaagaacctgacaaacaagatccaggcaataaagagcctga
aaaacaagaccctgttaaaatcaagatttattactctgataatgcaaccttaccatttaaagaagattggttagttataaaggaagctgagaa
gagatttaatgttgatttcgatttcgaagtaattccaattgcagattatcaaacaaaagtttctttaacattaaatacaggaaataacgctccagat
gtcatcctttatcagtcaacgcagggagagaatgcatct

Cald. kristjanssonii and *C. stercorarium* subs *leptospartum*

To the best of our knowledge, genome sequencing of the above organisms has not occurred and if it has, it has not been made available to the public. Based on our experimental results these organisms are cellulolytic and xylanolytic. The DNA sequences of genes encoding key metabolic enzymes are needed from these organisms in order to genetically engineer them and divert carbon flow to ethanol. These include such enzymes as acetate kinase and lactate dehydrogenase. In order to obtain the sequences of these genes, the genomes of these organisms will be sequenced.

With access to genome sequences, the conserved nature of the above enzymes may be used to find the encoding genes and flanking DNA. These sequences will be used to design constructs for targeted mutagenesis employing both single and double crossover strategies. These strategies will be identical to those described above. We will also determine which antibiotics can be used as selectable markers in these organisms and which protocols for transformation work best.

EXAMPLE 3

Transformation of *C. cellulolyticum*

Cells were grown in 50 mL of GS media with 4g/l cellobiose to an OD of 0.8 in anaerobic conditions, incubated at 34 degrees C. After harvesting they were washed 3 times in equal volumes with a wash buffer containing 500mM sucrose and 5mM MOPS with pH adjusted to 7. After the final wash, the cell pellet was resuspended in an equal volume of wash buffer 10ul aliquots of the cell suspension were placed in a standard electroporation cuvette with a 1mm electrode spacing. 1ul plasmid DNA was added. The concentration of the plasmid DNA was adjusted to ensure between a 1:1 and 10:1 molar ratio of plasmid to cells. A 5ms pulse was applied with a field strength of 7kV/cm (measured) across the sample. A custom pulse generator was used. The sample was immediately diluted 1000:1 with the same media used in the initial culturing and allowed to recover until growth resumed, and was determined via an increase in the OD (24-48h). The recovered sample was diluted 50:1 and placed in selective media with either 15ug/mL erythromycin or 15ug/mL chloramphenicol and allowed to grow for 5-6 days. Samples exhibiting growth in selective media were tested to confirm that they were in fact *C. cellulolyticum* and that they had the plasmid.

EXAMPLE 4

Constructs for Engineering Cellulolytic Strains

Cellulose is one of the main components of biomass, which can be potentially used as a substrate for generation of fuel ethanol by fermentation with *Clostridium thermocellum*. However, in this process, much energy and carbon sources are used to form by-product acetate and lactate. Engineering of the metabolic pathways of cellulose

utilization in *Clostridium thermocellum* is necessary to minimize the lactate and acetate production and make energy and carbon flows favorable to ethanol formation.

Acetate kinase is an important enzyme in the metabolic pathway of cellulose utilization to form acetate in *Clostridium thermocellum*, which is encoded by the *ack* gene. Inactivation of the *ack* gene may interrupt acetate kinase, leading to reduction or elimination of acetate.

Lactate dehydrogenase is an important enzyme in the metabolic pathway of cellulose utilization to form lactate in *Clostridium thermocellum*, which is encoded by the *ldh* gene. Inactivation of the *ldh* gene may interrupt lactate dehydrogenase, leading to reduction or elimination of lactate generation.

Inactivation of the ack Gene in C. thermocellum based on the plasmid pIKM1

To knock out the *ack* gene, a vector is constructed on the multiple cloning sites (MCS) of the plasmid pIKM1, in which the *cat* gene, encoding chloramphenicol acetyltransferase, is inserted into a DNA fragment of 3055 bp, involving the *ack* and the *pta* genes (encoding phosphotransacetylase), leading to knockout of 476 bp of the *ack* gene and 399 bp of the *pta* gene, and forming 1025 bp and 1048 bp flanking regions on both sides of the *mLs* gene respectively (Figure 7). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and *Staphylococcus aureus*, and is anticipated to be stably replicated in Gram positive strains of bacteria, including *Clostridium thermocellum*. The sequence of the ack knockout vector constructed on plasmid pIKM1 is set forth as SEQ ID NO:1.

Inactivation of the ack Gene in C. thermocellum based on the replicative plasmid pNW33N

To knock out the *ack* gene, a vector is constructed on the multiple cloning sites (MCS) of the replicative plasmid pNW33N, in which the macrolide, lincosamide, and streptogramin B (MLS_B) resistant gene *mLs* is inserted into a DNA fragment of 3345 bp, which includes the *ack* gene, the *pta* gene (encoding phosphotransacetylase) and an unknown upstream gene, leading to knockout of 855 bp of the *ack* gene and formation of flanking regions of 1195 bp and 1301 bp on either side of the *mLs* gene (Figure 8). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and *Staphylococcus aureus*, and is anticipated to be stably replicated in Gram positive strains of

bacteria, including *Clostridium thermocellum*. The sequence of the *ack* knockout vector constructed on plasmid pNW33N is set forth as SEQ ID NO:2.

Inactivation of the ldh Gene in C. thermocellum based on the plasmid pIKM1

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the plasmid pIKM1, in which the *cat* gene, encoding chloramphenicol acetyltransferase, is inserted into a DNA fragment of 3188 bp, involving the *ldh* and the *mdh* gene (encoding malate dehydrogenase), leading to knockout of a DNA fragment of 1171 bp, including part of the *ldh* and *mdh* genes, and forming 894 bp and 1123 bp flanking regions on both sides of the *mLs* gene, respectively (Figure 9). The sequence of the *ldh* knockout vector constructed on plasmid pIKM1 is set forth as SEQ ID NO:3.

Inactivation of the ldh Gene in C. thermocellum based on plasmid pNW33N

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the replicative plasmid pNW33N, in which the macrolide, lincosamide, and streptogramin B (MLS_B) resistant gene *mLs* is inserted into a DNA fragment of 2523 bp, which includes the *ldh* gene and the *mdh* gene (encoding malate dehydrogenase), leading to knocking out of a fragment of 489 bp of the *ldh* gene and formation of flanking regions of 1034 bp and 1000 bp on either side of the *mLs* gene (Figure 10). pNW33N contains pBC1 replicon, which is isolated from *Bacillus coagulans* and *Staphylococcus aureus*, and is anticipated to be stably replicated in other Gram positive strains of bacteria, including *Clostridium thermocellum*. The sequence of the *ldh* knockout vector constructed on plasmid pNW33N is set forth as SEQ ID NO:4.

Inactivation of the ldh Gene in Clostridium thermocellum based on plasmid pUC19

To knock out the *ldh* gene, a vector is constructed on the multiple cloning sites (MCS) of the pUC19 plasmid, in which a gene encoding chloramphenicol acetyltransferase (the *cat* gene) is inserted into a *ldh* gene fragment of 717 bp, leading to a flanking region of 245 bp and 255 bp on either side of the *cat* gene (Figure 11). pUC19 is an *E. coli* plasmid vector, containing pMB1 origin, which cannot be amplified in Gram positive strains of bacteria, including *Clostridium thermocellum*. A similar vector may be constructed, in which the mL*s* gene is flanked by the *ldh* gene fragments. The sequence of the *ldh* knockout vector constructed on plasmid pUC19 is set forth as SEQ ID NO:5.

Expression of xylose isomerase and xylulose kinase in C. thermocellum and C. straminisolvens (prophetic example)

For expression of *xylose isomerase* and *xylulose kinase* in *C. thermocellum*, the *xylose isomerase* and *xylulose kinase* genes were cloned from *T. saccharolyticum* and placed under control of the *C. thermocellum gapDH* promoter. This cassette is harbored in a *C. thermocellum* replicative plasmid based on the pNW33N backbone, resulting in pMU340 (Fig. 35) SEQ ID NO:74. Upon transfer into *C. thermocellum,* the resulting transformation can be assayed for the ability to grow on xylose. Analogous constructs can be created using the *C. kristajanssonii xylose isomerase* and *xylulose kinase* genes. These constructs can be tested for functionality in *C. straminsolvens* as well.

Expression of pyruvate decarboxylase and alcohol dehydrogenase in C. thermocellum and C. straminisolvens (prophetic example)

For expression of *pyruvate decarboxylase* and *alcohol dehydrogenase* in *C. thermocellum*, the *pyruvate decarboxylase* genes are cloned from sources *Z. mobilis* and *Z. palmae* and the *alcohol dehydrogenase* gene is cloned from source *Z. mobilis*. These genes (*pdc* and *adh*) will be expressed as an operon from the *C. thermocellum pta-ack* promoter. This cassette is harbored in a *C. thermocellum* replicative plasmid based on the pNW33N backbone (Figures 36 and 37), SEQ ID NOS:75 and 76. Upon transfer into *C. thermocellum,* the resulting transformation can be screened for enhanced ethanol production and/or aldehyde production to measure the functionality of the expressed enzymes. These constructs will be tested for functionality in *C. straminsolvens* as well.

EXAMPLE 5

Fermentation of Avicel® using *C. straminisolvens*

C. straminisolvens was used to ferment 1% Avicel® in serum bottles containing CTFUD medium. The product concentration profile and the ratios are shown in Figure 27. About 2 g/L of total products was generated in 3 d with ethanol constituting about 50% of the total products. Figure 27 shows the product concentration profiles for 1% Avicel® using *C. straminisolvens*. The ethanol to acetate ratio is depicted as E/A and the ratio of ethanol to total products is depicted as E/T.

EXAMPLE 6

Engineered Group II Introns For Mesophilic and Thermophilic Cellulolytic, Xylanolytic Organisms

Mobile group II introns, found in many bacterial genomes, are both catalytic RNAs and retrotransposable elements. They use a mobility mechanism known as retrotransposition in which the excised intron RNA reverse splices directly into a DNA target site and is then reverse transcribed by an intron-encoded protein. The mobile *Lactococcus lactis* Ll.LtrB group II intron has been developed into genetic tools known as Targetron™ vectors, which are commercially available from Sigma Aldritch (Catalog # TA0100). This product and its use are the subject of one or more of U.S. Patent Nos. 5,698,421, 5,804,418, 5,869,634, 6,027,895, 6,001,608, and 6,306,596 and/or other pending U.S. and foreign patent applications controlled by InGex, LLC.

Targetrons cassettes (Figures 28 and 29) which contain all the necessary sequences for retro-transposition may be sub-cloned into vectors capable of replication in mesophilic or thermophilic cellulolytic organisms. The Targetron cassette may be modified by replacing the lac promoter with any host- or species-specific constitutive or inducible promoters. The cassettes may be further modified through site-directed mutagenesis of the native recognition sequences such that the Group II intron is retargeted to insert into genes of interest creating genetic knockouts. For example, the group II intron could be redesigned to knockout lactate dehydrogenase or acetate kinase in any mesophilic or thermophilic cellulolytic organism. Table 4 depicts an example of insertion location and primers to retarget Intron to *C. cellulolyticum* acetate kinase (SEQ ID NO:21). Table 5 depicts an example of insertion location and primers to retarget Intron to *C. cellulolyticum* lactate dehydrogenase (SEQ ID NO:21).

An example of a vector for retargeting the L1.Ltrb intron to insert in *C. cell.* ack gene (SEQ ID NO:21) is depicted in Figure 28. The vector sequence of pMU367 (*C. cell.* acetate kinase KO vector) is SEQ ID NO:30.

An example of a vector for retargeting the L1.Ltrb intron to insert in *C. cell.* LDH2744 gene (SEQ ID NO:23) is depicted in Figure 29. The vector sequence of pMU367 (*C. cell.* lactate dehydrogenase KO vector) is set for as SEQ ID NO:31.

Table 4

Predicted Insertion location (SEQ ID NO:62)	ATTTACCTGGCTGGGAATACTGAGACATAT - intron - GTCATTGAGGCCGTA
IBS1 mutagenic primer (SEQ ID NO:63)	AAAAAAGCTTATAATTATCCTTA**ATTTCCTACTAC**GTGCGCCCAGATAGGGTG
EBS1d mutagenic primer (SEQ ID NO:64)	CAGATTGTACAAATGTGGTGATAACAGATAAGTC**TACTACTG**TAACTTACCTTTCTTTGT
EBS2 mutagenic primer (SEQ ID NO:65)	TGAACGCAAGTTTCTAATTTCG**GTTGAAATC**CGATAGAGGAAAGTGTCT

Table 5

Predicted Insertion location (SEQ ID NO:66)	TTAAATGTTGATAAGGAAGCTCTTTTCAAT - intron - GAAGTTAAGGTAGCA
IBS1 mutagenic primer (SEQ ID NO:67)	AAAAAAGCTTATAATTATCCTTA**GCTCTCTTCAAT**GTGCGCCCAGATAGGGTG
EBS1d mutagenic primer (SEQ ID NO 68)	CAGATTGTACAAATGTGGTGATAACAGATAAGTC**TTCAATGA**TAACTTACCTTTCTTTGT
EBS2 mutagenic primer (SEQ ID NO:69)	TGAACGCAAGTTTCTAATTTCG**ATTAGAGCT**CGATAGAGGAAAGTGTCT

EXAMPLE 7

Transformation of *Thermoanaerobacter* and *Thermoanaerobacterium strains (prophetic example)*

Thermoanaerobacter pseudoethanolicus 39E, *Thermoanaerobacterium saccharolyticum* JW/SL-YS485, *Thermoanaerobacterium saccharolyticum* B6A-RI, and *Thermoanaerobacter sp.* strain 59 will be transformed with the following protocol. Cells are grown at 55 °C in 40 mL of DSMZ M122 media (http://www.dsmz.de/microorganisms/media_list.php) with the following modifications: 5 g/L cellobiose instead of cellulose, 1.8 g/L K_2HPO_4, no glutathione, and 0.5 g/L L-cystiene-

HCl until an optical density of 0.6 to 0.8. Cells are then harvested and washed twice with 40 mL 0.2 M cellobiose at room temperature. Cells are re-suspended in 0.2 M cellobiose in aquilots of 100 uL and 0.1 to 1 ug plasmid DNA is added to the sample in a 1 mm gap-width electroportation cuvette. An exponential pulse (Bio-Rad Instruments) of 1.8 kV, 25 μF, 200Ω, ~ 3-6 ms is applied to the cuvette, and cells are diluted 100-200 fold in fresh M122 and incubated for 12-16 hours at 55°C. The recovered cells are then diluted 25-100 fold in petri-plates with fresh agar-containing media containing a selective agent, such as 200 μg/mL kanamycin. Once the media has solidified, plates incubated at 55°C for 24-72 hours for colony formation. Colonies can be tested by PCR for evidence of site-specific recombination.

EXAMPLE 8

Fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains

Table 6 depicts the fermentation performance of engineered *Thermoanaerobacter* and *Thermoanaerobacterium* strains. Cultures were grown for 24 hours in M122 at 55°C without shaking. The following abbreviations are used in Table 6: Cellobiose (CB), glucose (G), lactic acid (LA), acetic acid (AA), and ethanol (Etoh). Values are in grams per liter. YS485 – *Thermoanaerobacterium saccharolyticum* JW/SL-YS485, B6A-RI - *Thermoanaerobacterium saccharolyticum* B6A-RI, 39E – *Thermoanaerobacter pseudoethanolicus* 39E.

Table 6

Fermentation sample	CB	G	LA	AA	Etoh
YS485 wildtype	0	0	0.77	1.04	1.40
YS485 ΔL-ldh	0	0	0	0.92	1.73
YS485 Δpta/ack	2.51	0	0.75	0.06	0.62
YS485 ΔL-ldh, Δpta/ack	0	0	0	0	2.69
B6A-RI wildtype	0	0	0	1.0	1.76
B6A-RI ΔL-ldh, Δpta/ack strain #1	0	0	0	0	2.72
B6A-RI ΔL-ldh, Δpta/ack strain #2	0.45	0	0	0	2.49
39E wildtype	0.51	0	1.51	0.15	1.87
Media	5.10	0.25	0	0	0

EXAMPLE 9

Construct for Engineering Cellulolytic and Xylanolytic Strains -

Antisense RNA technology example

A replicative plasmid (Figure 38) carrying an antisense RNA cassette targeting a *C. thermocellum* gene coding for lactate dehydrogenase (Cthe_1053) was transferred to *C. thermocellum* 1313 by electroporation and thiamphenicol selection. The transformation efficiency observed for this plasmid was equal to that of the parent vector, pMU102. The sequence of the plasmid is shown in SEQ ID NO: 61. The asRNA cassette is depicted in Figure 38 and is organized as follows: (i) the entire 1827 bp cassette is cloned into the multicloning site of pMU102 in the orientation shown in Figure 38, (ii) the native promoter region is contained within the first 600 bp of the cassette, (iii) the first 877 bp of the *ldh* open reading frame are fused to the native promoter in the antisense orientation, (iv) approximately 300 additional bp are included downstream of the asRNA *ldh* region.

The resulting thiamphenicol resistant colonies were screened for altered end product formation by growing standing cultures on M122C media in the presence of 6 ug/mL thiamphenicol (to maintain the plasmid), as shown in Figure 39. A preliminary screen of 9 randomly selected thiamphenicol-resistant transformants showed that 4 cultures exhibited low levels of lactate production relative to wild type. Additionally, a construct carrying antisense RNA directed to both *ldh* genes are to be constructed in order to partially, substantially, or completely delete, silence, inactivate, or down-regulate both genes simultaneously.

EXAMPLE 10

SEQ ID NOS:44, 45, and 46 are the pyruvate-formate-lyase (aka formate acetyltransferase, EC. 2.3.1.54, pfl) genes from *Thermoanaerobacterium saccharolyticum* YS485, *Clostridium thermocellum* ATCC 27405, and *Clostridium phytofermentans*. Pfl catalyzes the conversion of pyruvate to Acetyl-CoA and formate (Figure 34). Deletion of pfl will result in the elimination of formate production, and could result in a decrease in acetic acid yield in some thermophilic strains, with a resulting increase in ethanol yield.

SEQ ID NOS:47-52, depicted in Figures 40-45, show pfl knockout plasmids, two each for the three organisms listed above. Each organism has a single crossover and double crossover plasmid designed to partially, substantially, or completely delete, silence, inactivate, or down-regulate the pfl enzyme. Single crossover plasmids are designed with a single DNA sequence (400 bp to 1000 bp) homologous to an internal section of the pfl gene, double crossover plasmids are designed with two DNA sequences (400 to 1000 bp) homologous to regions upstream (5') and downstream (3') to the pfl gene. All plasmids are designed to use the best available antibiotic markers for selection in the given organism. Plasmids can be maintained in *E. coli* and constructed through a DNA synthesis contract company, such as Codon Devices or DNA 2.0.

INCORPORATION BY REFERENCE

All of the U.S. patents and U.S. published patent applications cited herein are hereby incorporated by reference.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

1. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, or a complement thereof.

2. A genetic construct comprising SEQ ID NO:1 operably linked to a promoter expressible in a thermophilic or mesophilic bacterium.

3. A recombinant thermophilic or mesophilic bacterium comprising the genetic construct of claim 2.

4. A vector comprising the nucleic acid molecule of claim 1.

5. A host cell comprising the nucleic acid molecule of claim 1.

6. The host cell of claim 5, wherein the host cell is a thermophilic or mesophilic bacterial cell.

7. A genetically modified thermophilic or mesophilic microorganism, wherein the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1, thereby increasing the native ability of said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

8. The genetically modified microorganism according to claim 7, wherein said microorganism is a Gram-negative bacterium or a Gram-positive bacterium.

9. The genetically modified microorganism according to claim 7, wherein said microorganism is a species of the genera *Thermoanaerobacterium*, *Thermoanaerobacter*, *Clostridium*, *Geobacillus*, *Saccharococcus*, *Paenibacillus*, *Bacillus*, *Caldicellulosiruptor*, *Anaerocellum*, or *Anoxybacillus*.

10. The genetically modified microorganism according to claim 7, wherein said microorganism is a bacterium selected from the group consisting of: *Thermoanaerobacterium thermosulfurigenes*, *Thermoanaerobacterium aotearoense*, *Thermoanaerobacterium polysaccharolyticum*, *Thermoanaerobacterium zeae*, *Thermoanaerobacterium xylanolyticum*,

Thermoanaerobacterium saccharolyticum, *Thermoanaerobium brockii*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter thermohydrosulfuricus*, *Thermoanaerobacter ethanolicus*, *Thermoanaerobacter brocki*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium phytofermentans*, *Clostridium straminosolvens*, *Geobacillus thermoglucosidasius*, *Geobacillus stearothermophilus*, *Saccharococcus caldoxylosilyticus*, *Saccharoccus thermophilus*, *Paenibacillus campinasensis*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor acetigenus*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, and *Anaerocellum thermophilum*.

11. The genetically modified microorganism according to claim 7, wherein said microorganism is *Thermoanaerobacterium saccharolyticum*.

12. The genetically modified microorganism according to claim 7, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic microorganism with a native ability to metabolize a hexose sugar; (b) a thermophilic or mesophilic microorganism with a native ability to metabolize a pentose sugar; and (c) a thermophilic or mesophilic microorganism with a native ability to metabolize a hexose sugar and a pentose sugar.

13. The genetically modified microorganism according to claim 7, wherein said microorganism has a native ability to metabolize a hexose sugar.

14. The genetically modified microorganism according to claim 13, wherein said microorganism is *Clostridium straminisolvens* or *Clostridium thermocellum*.

15. The genetically modified microorganism according to claim 7, wherein said microorganism has a native ability to metabolize a hexose sugar and a pentose sugar.

16. The genetically modified microorganism according to claim 15, wherein said microorganism is *Clostridium cellulolyticum*, *Clostridium kristjanssonii*, or *Clostridium stercorarium subsp. leptosaprartum*.

17. The genetically modified microorganism of claim 13, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a pentose sugar.

18. The genetically modified microorganism according to claim 7, wherein said microorganism has a native ability to metabolize a pentose sugar.

19. The genetically modified microorganism according to claim 18, wherein said microorganism is selected from the group consisting of *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium polysaccharolyticum*, and *Thermoanaerobacterium thermosaccharolyticum*.

20. The genetically modified microorganism of claim 18, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product from a hexose sugar.

21. The genetically modified microorganism according to any one of claims 7-20, wherein a native gene is partially, substantially, or completely deleted, silenced, inactivated, or down-regulated, which native gene encodes a native enzyme involved in the metabolic production of an organic acid or a salt thereof.

22. The genetically modified microorganism according to claim 21, wherein said native enzyme is phosphotransacetylase.

23. The genetically modified microorganism according to claim 21, wherein said native enzyme is lactate dehydrogenase.

24. A genetically modified thermophilic or mesophilic microorganism, wherein (a) the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1; and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme

involved in the metabolic production of ethanol, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

25. The genetically modified microorganism according to claim 24, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar.

26. The genetically modified microorganism according to claim 24, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a pentose sugar.

27. The genetically modified microorganism according to claim 24, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to metabolize a hexose sugar; and a second non-native gene is inserted, which second non-native gene encodes a second non-native enzyme that confers the ability to metabolize a pentose sugar, thereby allowing said thermophilic or mesophilic microorganism to metabolize a hexose sugar and a pentose sugar.

28. The genetically modified microorganism according to any one of claims 24, 25 or 27, wherein said first non-native enzyme is pyruvate decarboxylase (PDC) or alcohol dehydrogenase (ADH).

29. The genetically modified microorganism according to claim 24 or claim 26, wherein said first non-native enzyme is xylose isomerase.

30. The genetically modified microorganism according to claim 29, wherein said first non-native gene corresponds to SEQ ID NOS:6, 10 or 14.

31. The genetically modified microorganism according to any one of claims 24, 26 or 27, wherein said non-native enzyme is xylulokinase.

32. The genetically modified microorganism according to claim 31, wherein said non-native gene corresponds to SEQ ID NOS:7, 11 or 15.

33. The genetically modified microorganism according to any one of claims 24, 26 or 27, wherein said non-native enzyme is L-arabinose isomerase.

34. The genetically modified microorganism according to claim 33, wherein said non-native gene corresponds to SEQ ID NOS:8 or 12.

35. The genetically modified microorganism according to any one of claims 24, 26 or 27, wherein said non-native enzyme is L-ribulose-5-phosphate 4-epimerase.

36. The genetically modified microorganism according to claim 35, wherein said non-native gene corresponds to SEQ ID NOS: 9 or 13.

37. The genetically modified microorganism according to any one of claims 7-27 or 28-36, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol.

38. The genetically modified microorganism according to claim 7, wherein said microorganism is selected from the group consisting of: (a) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose; (b) a thermophilic or mesophilic microorganism with a native ability to hydrolyze xylan; and (c) a thermophilic or mesophilic microorganism with a native ability to hydrolyze cellulose and xylan.

39. The genetically modified microorganism according to claim 7 or claim 24, wherein said microorganism has a native ability to hydrolyze cellulose.

40. The genetically modified microorganism according to claim 7 or claim 24, wherein said microorganism has a native ability to hydrolyze cellulose and xylan.

41. The genetically modified microorganism according to claim 39, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan.

42. The genetically modified microorganism according to claim 7 or claim 24, wherein said microorganism has a native ability to hydrolyze xylan.

43. The genetically modified microorganism according to claim 42, wherein a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose.

44. The genetically modified microorganism according to any one of claims 39-43, wherein a native gene is partially, substantially, or completely-deleted, silenced, inactivated, or down-regulated, which native gene encodes a native enzyme involved in the metabolic production of an organic acid or a salt thereof.

45. The genetically modified microorganism according to claim 44, wherein said native enzyme is phosphotransacetylase.

46. The genetically modified microorganism according to claim 44, wherein said native enzyme is lactate dehydrogenase.

47. A genetically modified thermophilic or mesophilic microorganism, wherein (a) the genetically modified thermophilic or mesophilic microorganism has been transformed by a nucleotide sequence of SEQ ID NO:1; and (b) a first non-native gene is inserted, which first non-native gene encodes a first non-native enzyme involved in the hydrolysis of a polysaccharide, thereby allowing said thermophilic or mesophilic microorganism to produce ethanol as a fermentation product.

48. The genetically modified microorganism according to claim 47, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose.

49. The genetically modified microorganism according to claim 47, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze xylan.

50. The genetically modified microorganism according to claim 47, wherein said first non-native gene encodes a first non-native enzyme that confers the ability to hydrolyze cellulose; and a second non-native gene is inserted, which second non-

native gene encodes a second non-native enzyme that confers the ability to hydrolyze xylan, thereby allowing said thermophilic or mesophilic microorganism to hydrolyze cellulose and xylan.

51. The genetically modified microorganism according to any one of claims 47, 48 or 50, wherein said first non-native enzyme is PDC or ADH.

52. The genetically modified microorganism according to any one of claims 38-51, wherein said microorganism is able to convert at least 60% of carbon from metabolized biomass into ethanol.

53. The genetically-modified microorganism according to any one of claims 7, 9, 10, 12, 14, 16, 17, 20, 24-27, 38 or 47-50, wherein said microorganism is mesophilic.

54. The genetically-modified microorganism according to any one of claims 7, 9-12, 17, 19, 20, 24-27, 38 or 47-50, wherein said microorganism is thermophilic.

55. A process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to any one of claims 7-37.

56. The process of claim 55, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, and a combination thereof.

57. The process of claim 55, wherein said lignocellulosic biomass is selected from the group consisting of corn stover, sugarcane bagasse, switchgrass, and poplar wood.

58. The process of claim 55, wherein said lignocellulosic biomass is willow.

59. A process for converting lignocellulosic biomass to ethanol, comprising contacting lignocellulosic biomass with a genetically modified thermophilic or mesophilic microorganism according to any one of claims 38-54.

60. The process of claim 59, wherein said lignocellulosic biomass is selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, mixed prairie grass, miscanthus, sugar-processing residues, sugarcane bagasse, sugarcane straw agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, and a combination thereof.

61. The process of claim 59, wherein said lignocellulosic biomass is selected from the group consisting of corn stover, sugarcane bagasse, switchgrass, and poplar wood.

62. The process of claim 59, wherein said lignocellulosic biomass is willow.

63. Ethanol produced by the process according to any one of claims 55 to 62.

64. An isolated nucleic acid molecule according to claim 1; or a genetic construct according to claim 2; or a recombinant thermophilic or mesophilic bacterium according to claim 3; or a vector according to claim 4; or a host cell according to claim 5; or a genetically modified thermophilic or mesophilic microorganism according to any one of claims 7, 24 or 47; or a process for converting lignocellulosic biomass to ethanol according to claim 55 or claim 59; or ethanol produced by the process according to claim 55 or claim 59, substantially as herein described with reference to any one or more of the examples but excluding comparative examples.

Figure 1

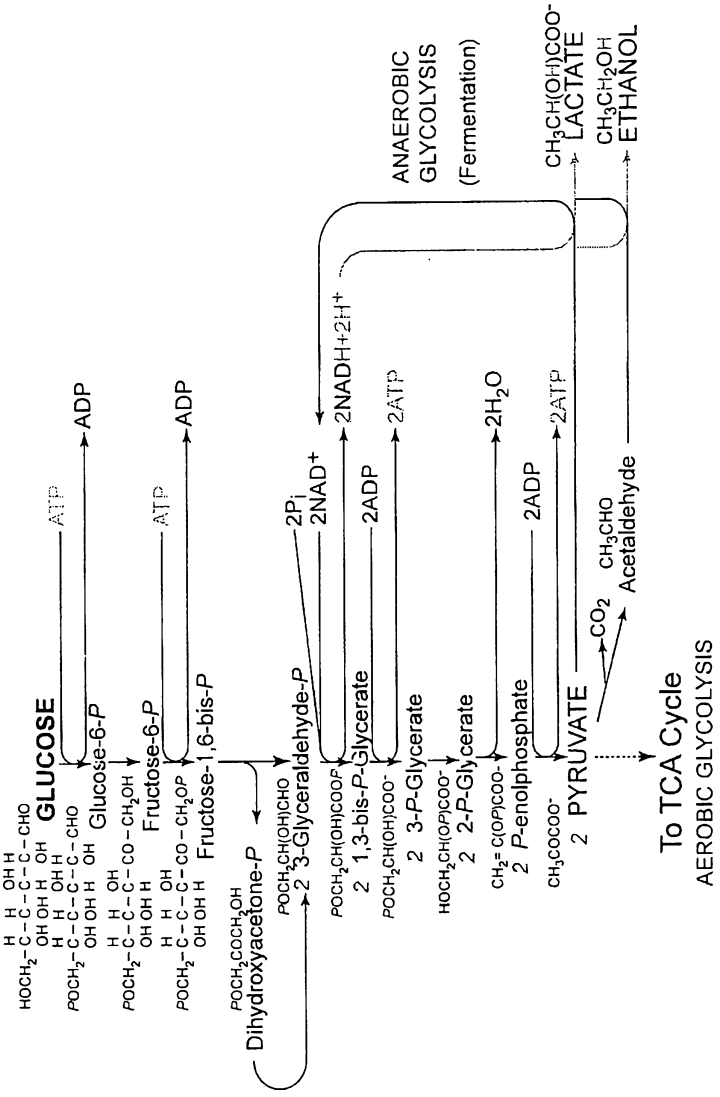
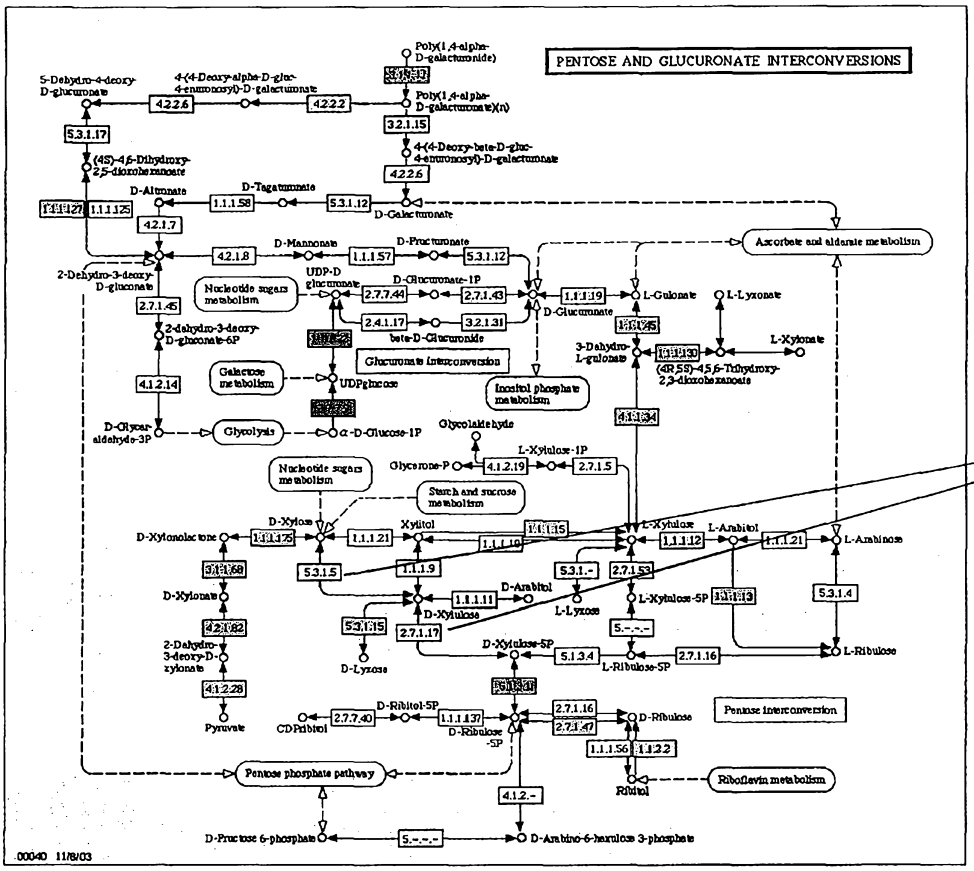

GLYCOLYSIS
GLUCOSE
ATP
ADP
Glucose-6-P
Fructose-6-P
ATP
ADP
Fructose-1,6-bis-P
Dihydroxyacetone-P
2 3-Glyceraldehyde-P
2Pi
2NAD+
2NADH+2H+
2 1,3-bis-P-Glycerate
2ADP
2ATP
2 3-P-Glycerate
2 2-P-Glycerate
2H2O
2 P-enolphosphate
2ADP
2ATP
2 PYRUVATE
CO2
CH3CHO
Acetaldehyde
To TCA Cycle
AEROBIC GLYCOLYSIS
ANAEROBIC
GLYCOLYSIS
(Fermentation)
CH3CH(OH)COO-
LACTATE
CH3CH2OH
ETHANOL

Figure 2



XI (5.3.1.5) and XK (2.7.1.17 - Missing

Figure 3

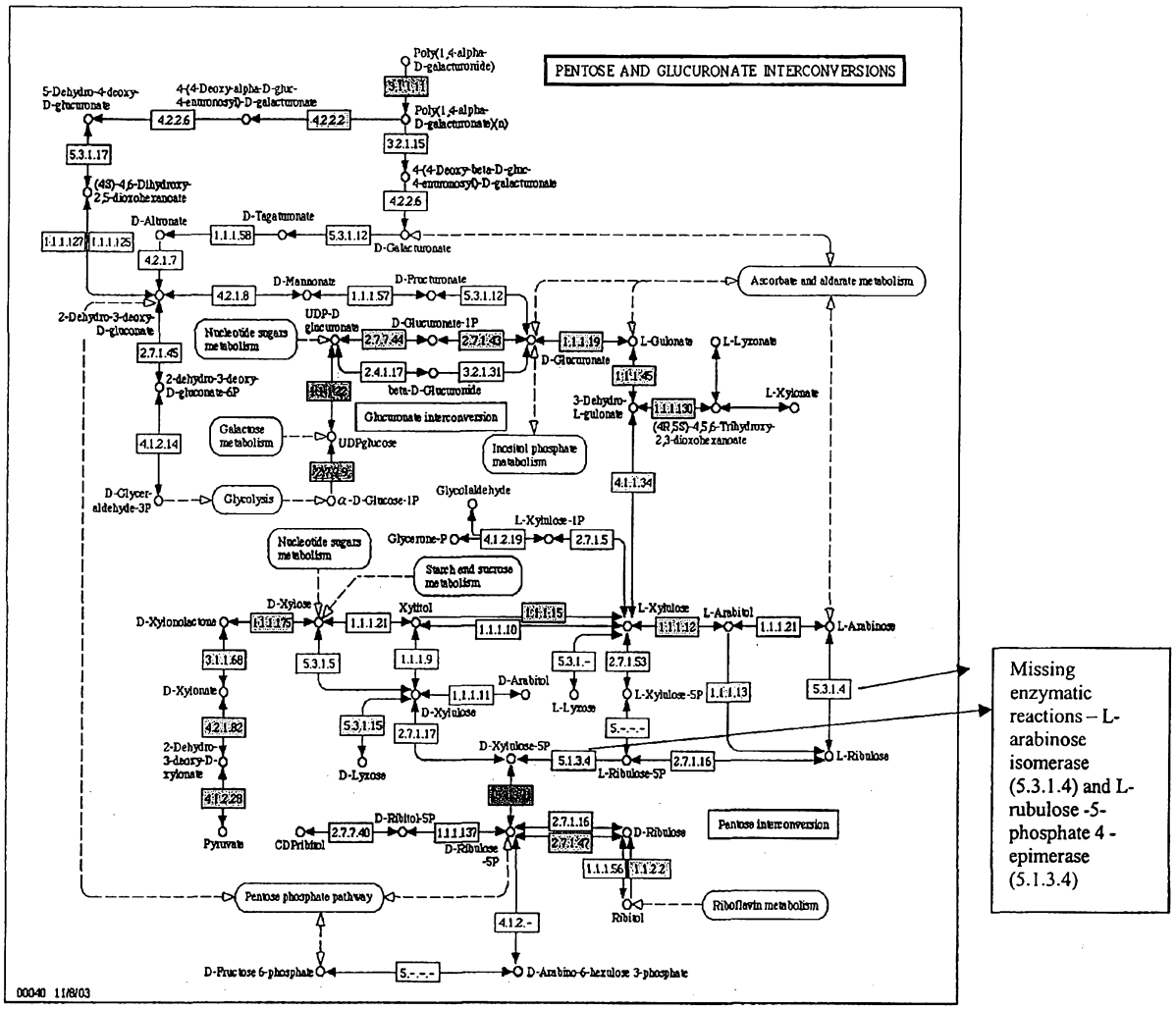

PENTOSE AND GLUCURONATE INTERCONVERSIONS
Poly(1,4-alpha-D-galacturonide)
5-Dehydro-4-deoxy-D-glucuronate
4-(4-Deoxy-alpha-D-gluc-4-enuronosyl)-D-galacturonate
Poly(1,4-alpha-D-galacturonate)(n)
4-(4-Deoxy-beta-D-gluc-4-enuronosyl)-D-galacturonate
(4S)-4,5-Dihydroxy-2,5-dioxohexanoate
D-Altronate
D-Tagaturonate
D-Galacturonate
D-Mannonate
D-Fructuronate
Ascorbate and aldarate metabolism
2-Dehydro-3-deoxy-D-gluconate
UDP-D glucuronate
D-Glucuronate-1P
D-Glucuronate
L-Gulonate
L-Lyxonate
2-dehydro-3-deoxy-D-gluconate-6P
beta-D-Glucuronide
3-Dehydro-L-gulonate
L-Xylonate
(4R,5S)-4,5,6-Trihydroxy-2,3-dioxohexanoate
Nucleotide sugars metabolism
Galactose metabolism
Glucuronate interconversion
UDPglucose
Inositol phosphate metabolism
D-Glycer-aldehyde-3P
Glycolysis
α-D-Glucose-1P
Glycolaldehyde
L-Xylulose-1P
Glycerone-P
Starch and sucrose metabolism
D-Xylose
Xylitol
L-Xylulose
L-Arabitol
L-Arabinose
D-Xylonolactone
D-Xylonate
D-Arabitol
L-Lyxose
L-Xylulose-5P
D-Xylulose
D-Lyxose
D-Xylulose-5P
L-Ribulose
L-Ribulose-5P
2-Dehydro-3-deoxy-D-xylonate
Pyruvate
D-Ribitol-5P
CDPribitol
D-Ribulose
D-Ribulose-5P
Pentose interconversion
Ribitol
Riboflavin metabolism
Pentose phosphate pathway
D-Fructose 6-phosphate
D-Arabino-6-hexulose 3-phosphate
00040 11/8/03
Missing enzymatic reactions – L-arabinose isomerase (5.3.1.4) and L-rubulose -5-phosphate 4 -epimerase (5.1.3.4)

Figure 4

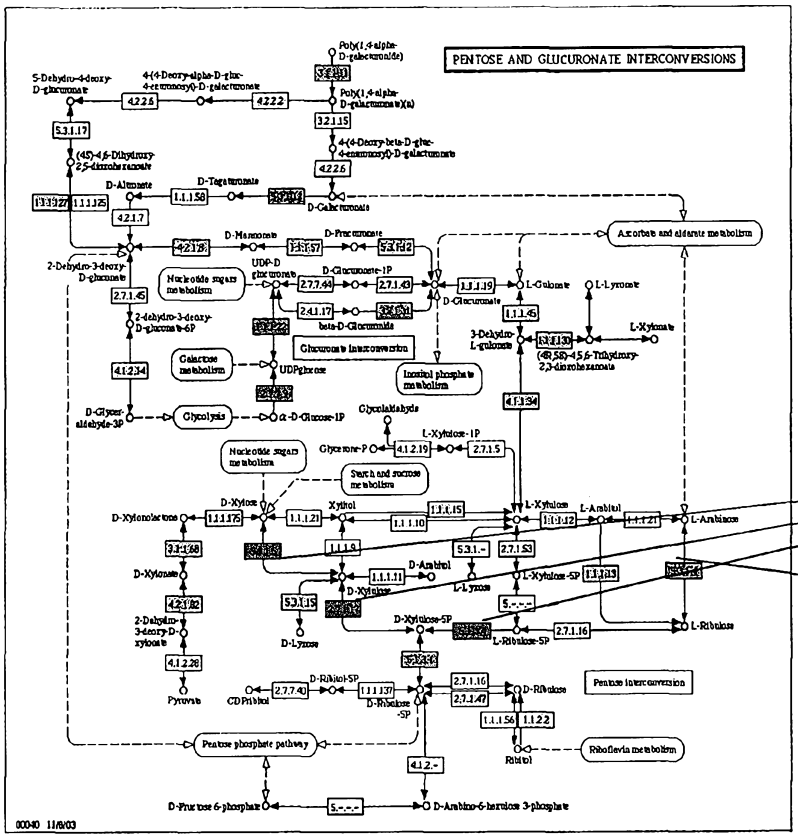

PENTOSE AND GLUCURONATE INTERCONVERSIONS
All four genes present

Figure 5

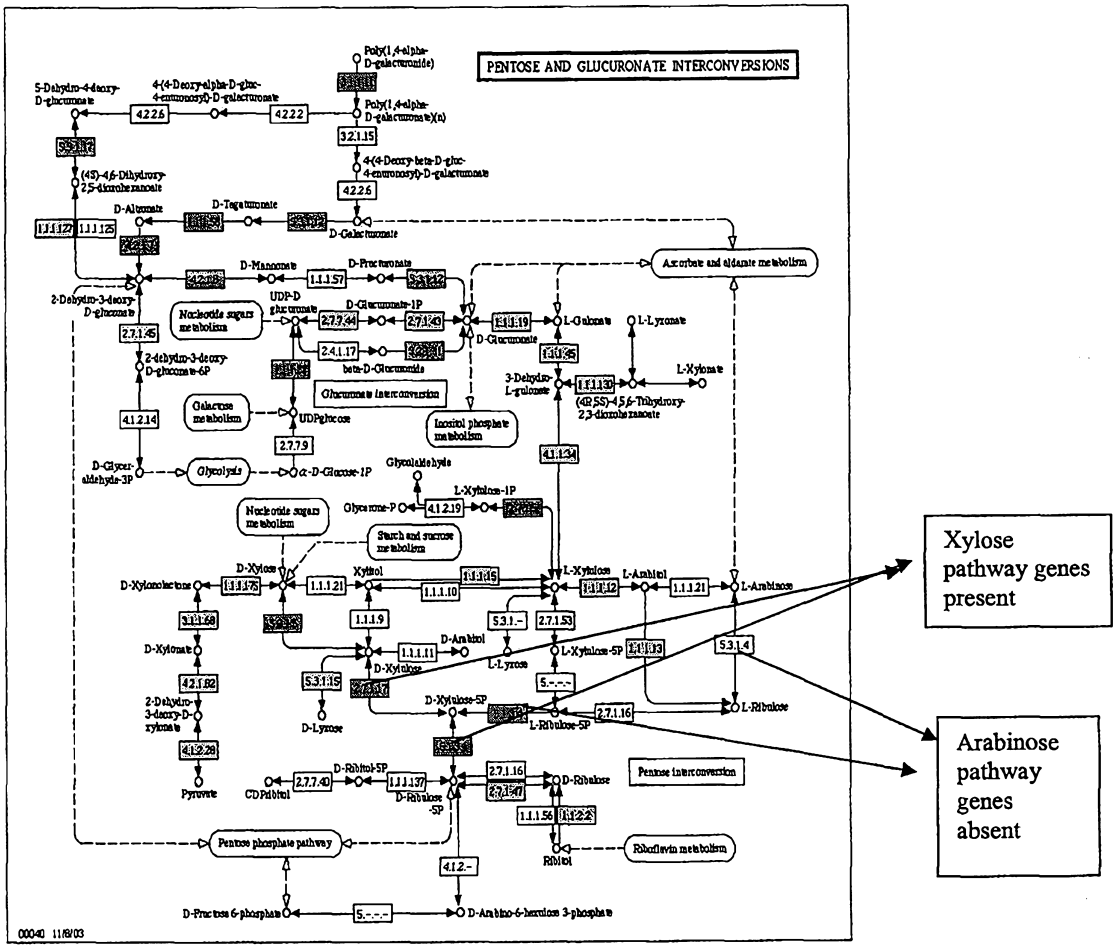

PENTOSE AND GLUCURONATE INTERCONVERSIONS
Xylose pathway genes present
Arabinose pathway genes absent

Figure 6

16 Apr 2007 Alignment Results

Alignment: Global DNA alignment against reference molecule
Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

Reference molecule: Clostridium thermocellum, Region 1 to 1509
Number of sequences to align: 7
Total length of aligned sequences with gaps: 1767 bps

Sequence	Start	End	#Match	NonMatch	%Match
Clostridium thermocellum	1	1509			
Clostridium cellulolyticum	1	1642	1372	274	83
Thermoanaerobacterium sacc	1	1552	1286	278	82
C. stercorarium	1	1519	1374	155	89
C. stercorarium II	1	1500	1339	194	87
Caldicellulosiruptor krist	1	1508	1250	300	80
C. phytofermentans	1	1371	1115	403	73

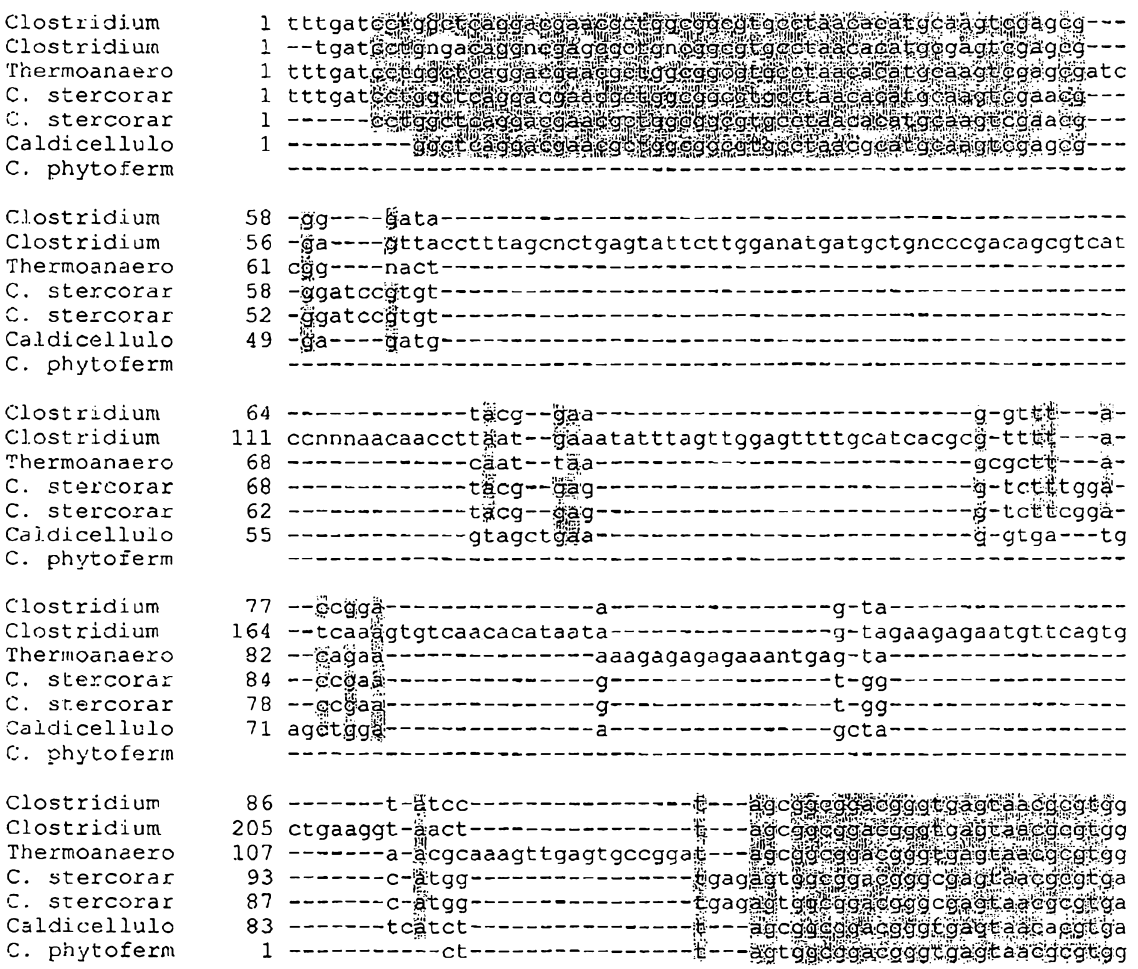
16 Apr 2007 Alignment Results

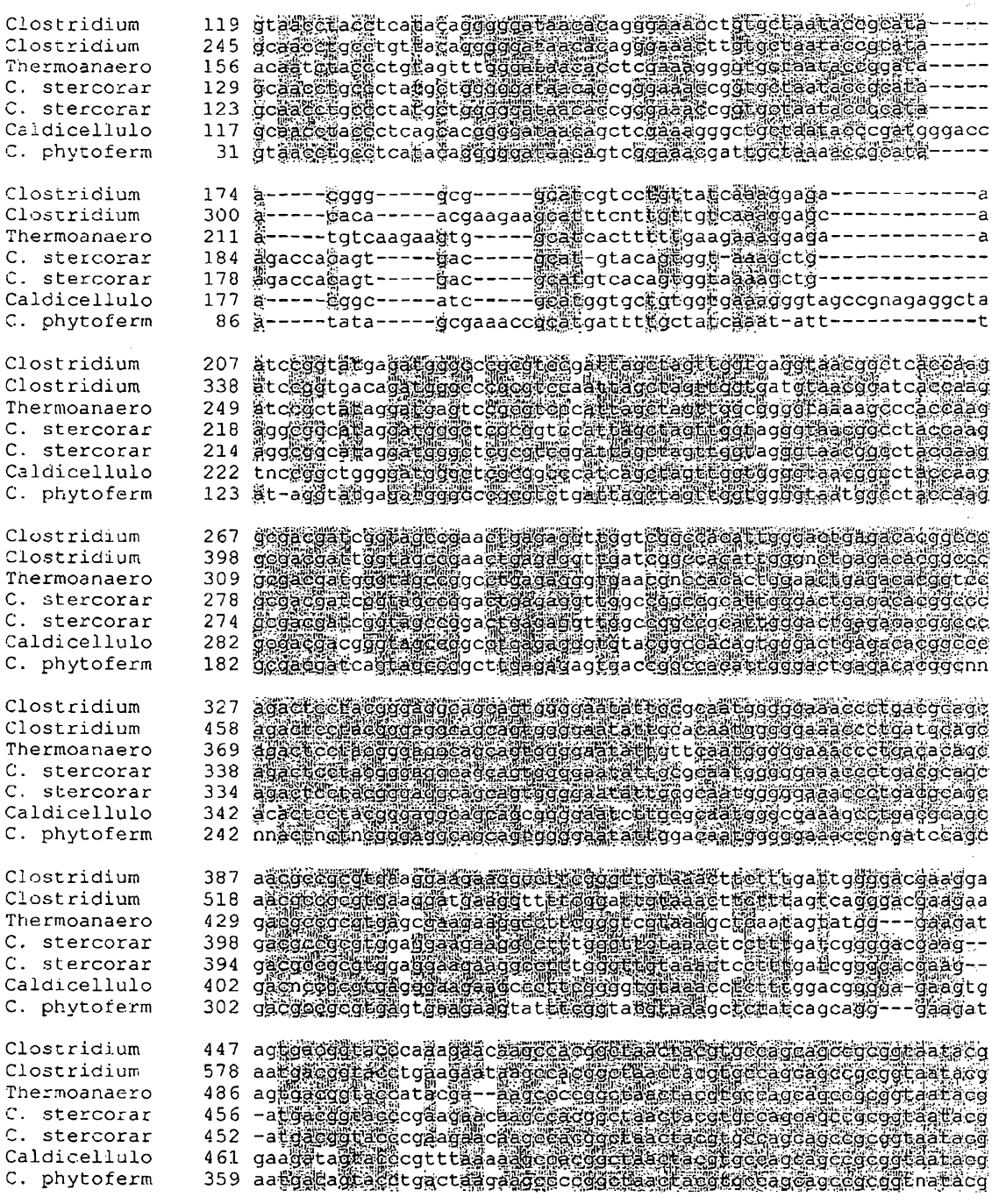
Alignment: Global DNA alignment against reference molecule
Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

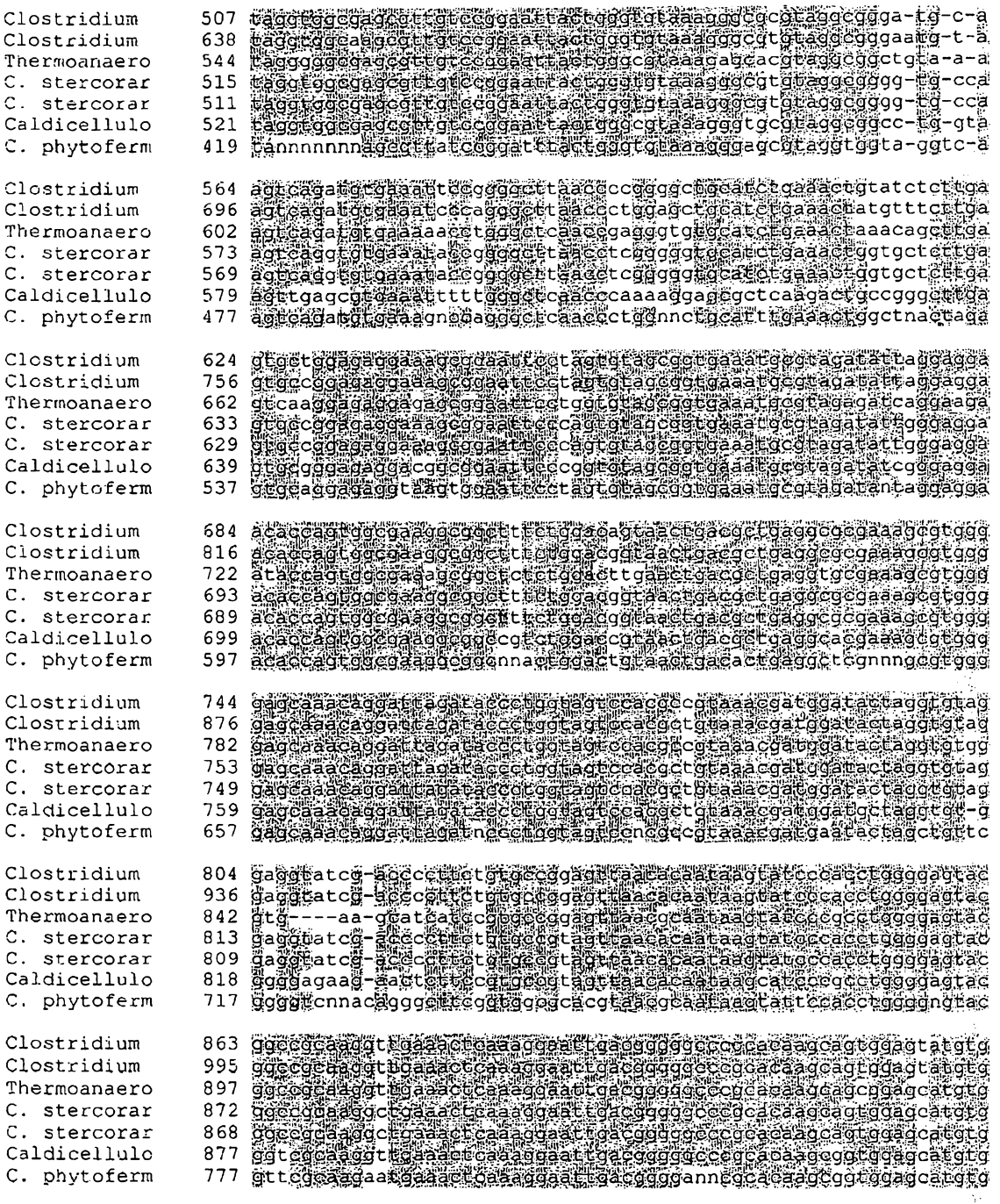
Reference molecule: Clostridium thermocellum, Region 1 to 1509
Number of sequences to align: 7
Total length of aligned sequences with gaps: 1767 bps
Settings: Similarity significance value cutoff: >= 60%

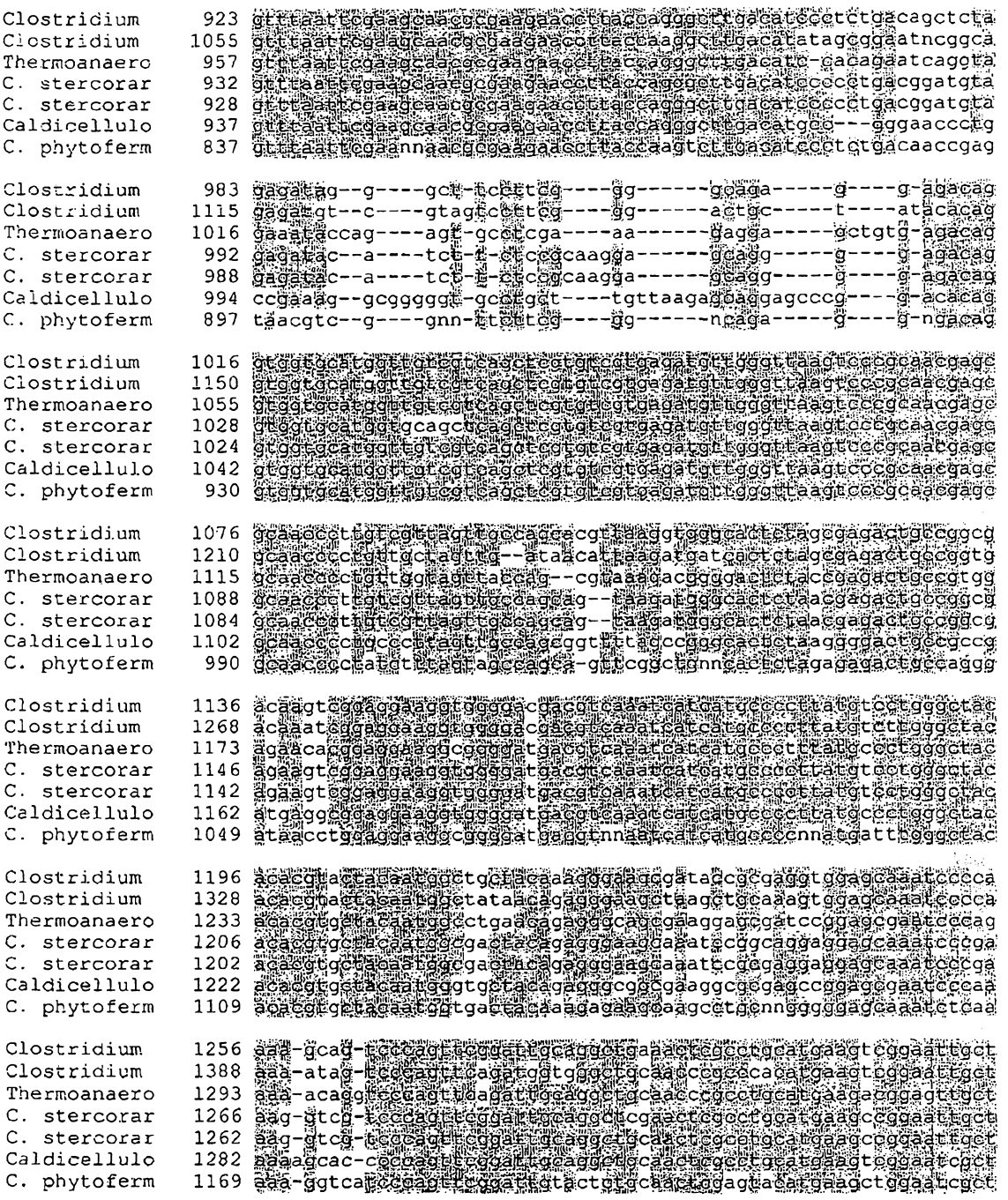
Summary of Percent Matches:

Ref:	Clostridium thermocellum	1 to 1509	(1509 bps)	--		
2:	Clostridium cellulolyticum	1 to 1642	(1642 bps)	83%		
3:	Thermoanaerobacterium saccharolyticum	1 to 1552	(1552 bps)	82%		
4:	C. stercorarium	1 to 1519	(1519 bps)	89%		
5:	C. stercorarium II	1 to 1500	(1500 bps)	87%		
6:	Caldicellulosiruptor kristjanssonii	1 to 1508	(1508 bps)	80%		
7:	C. phytofermentans	1 to 1371	(1371 bps)	73%		






Clostridium 507
Clostridium 638
Thermoanaero 544
C. stercorar 515
C. stercorar 511
Caldicellulo 521
C. phytoferm 419
Clostridium 564
Clostridium 696
Thermoanaero 602
C. stercorar 573
C. stercorar 569
Caldicellulo 579
C. phytoferm 477
Clostridium 624
Clostridium 756
Thermoanaero 662
C. stercorar 633
C. stercorar 629
Caldicellulo 639
C. phytoferm 537
Clostridium 684
Clostridium 816
Thermoanaero 722
C. stercorar 693
C. stercorar 689
Caldicellulo 699
C. phytoferm 597
Clostridium 744
Clostridium 876
Thermoanaero 782
C. stercorar 753
C. stercorar 749
Caldicellulo 759
C. phytoferm 657
Clostridium 804
Clostridium 936
Thermoanaero 842
C. stercorar 813
C. stercorar 809
Caldicellulo 818
C. phytoferm 717
Clostridium 863
Clostridium 995
Thermoanaero 897
C. stercorar 872
C. stercorar 868
Caldicellulo 877
C. phytoferm 777



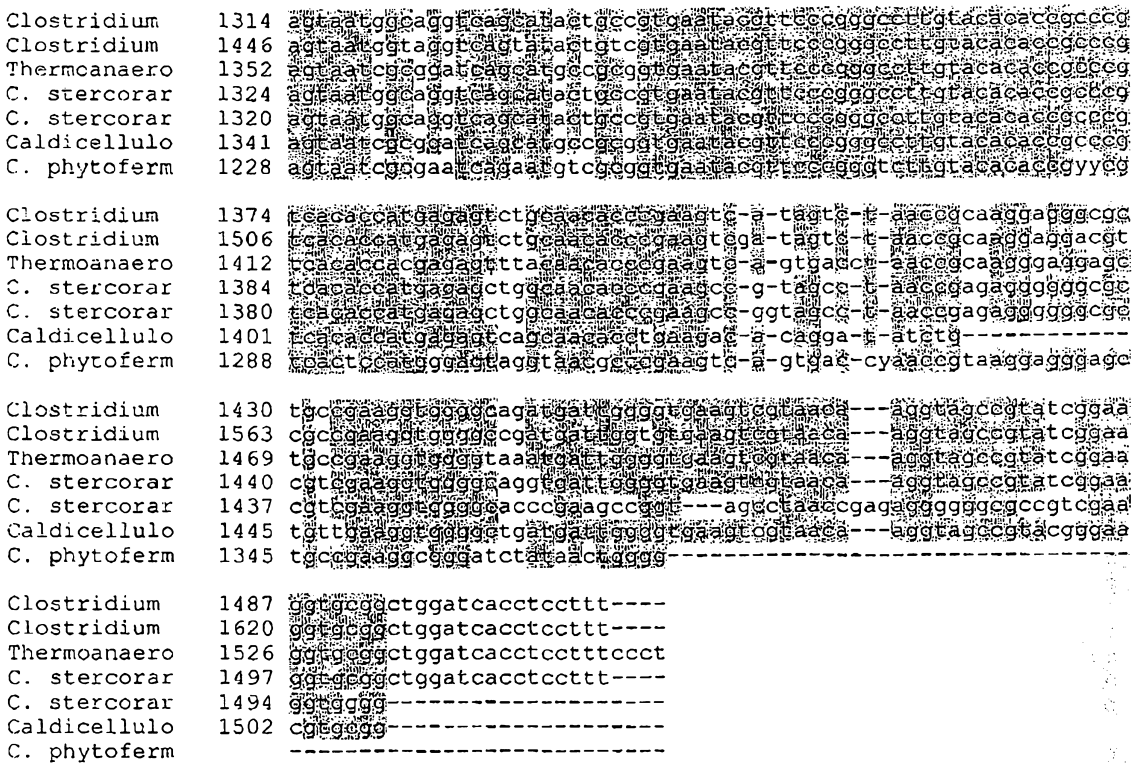

Clostridium 1314
Clostridium 1446
Thermoanaero 1352
C. stercorar 1324
C. stercorar 1320
Caldicellulo 1341
C. phytoferm 1228
Clostridium 1374
Clostridium 1506
Thermoanaero 1412
C. stercorar 1384
C. stercorar 1380
Caldicellulo 1401
C. phytoferm 1288
Clostridium 1430
Clostridium 1563
Thermoanaero 1469
C. stercorar 1440
C. stercorar 1437
Caldicellulo 1445
C. phytoferm 1345
Clostridium 1487
Clostridium 1620
Thermoanaero 1526
C. stercorar 1497
C. stercorar 1494
Caldicellulo 1502
C. phytoferm

Figure 7

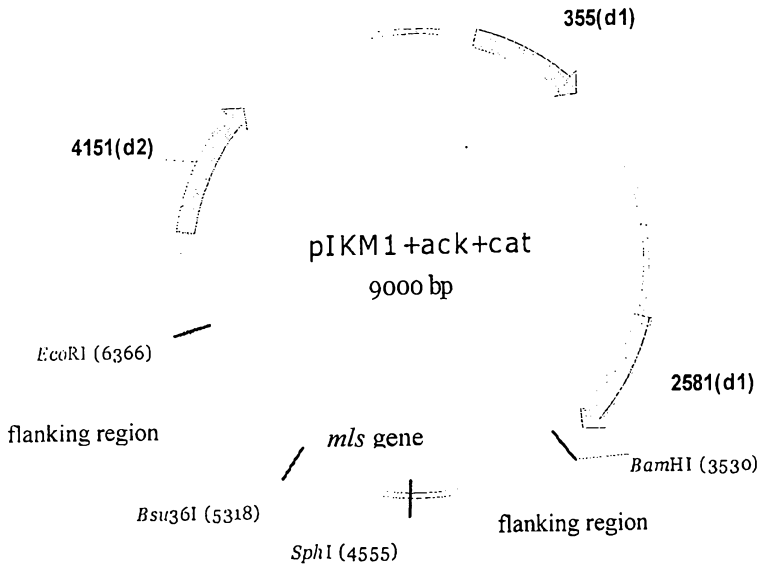

355(d1)
4151(d2)
pIKM1+ack+cat
9000 bp
EcoRI (6366)
2581(d1)
flanking region
mls gene
BamHI (3530)
Bsu36I (5318)
flanking region
SphI (4555)

Figure 8

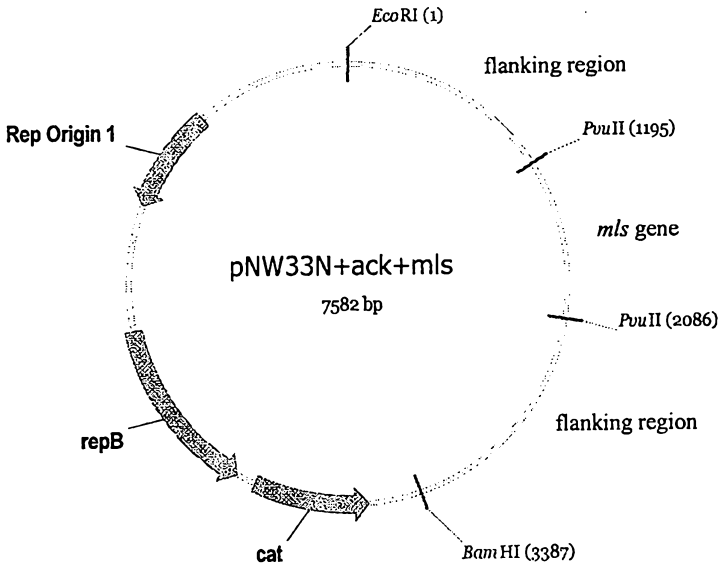

EcoRI (1)
flanking region
PvuII (1195)
Rep Origin 1
mls gene
pNW33N+ack+mls
7582 bp
PvuII (2086)
flanking region
repB
cat
BamHI (3387)

Figure 9

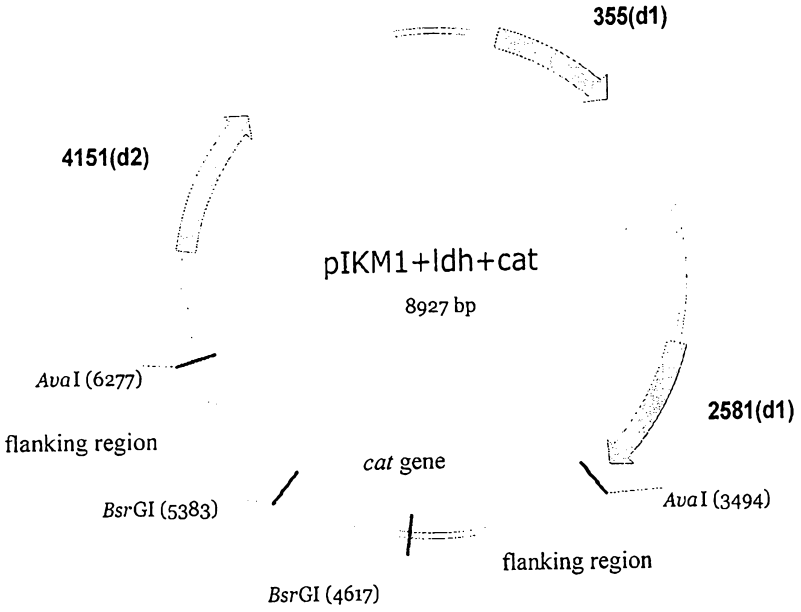

355(d1)
4151(d2)
pIKM1+ldh+cat
8927 bp
AvaI (6277)
flanking region
2581(d1)
cat gene
AvaI (3494)
BsrGI (5383)
flanking region
BsrGI (4617)

Figure 10

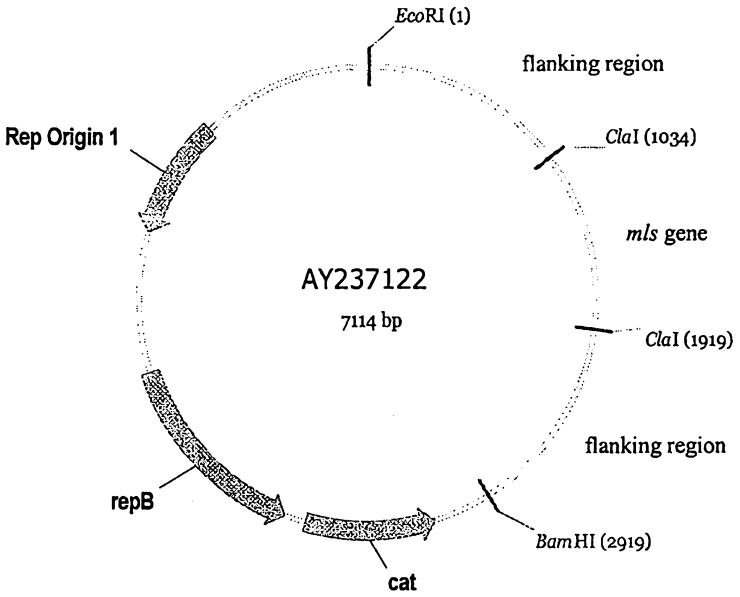

EcoRI (1)
flanking region
ClaI (1034)
mls gene
ClaI (1919)
flanking region
BamHI (2919)
cat
repB
Rep Origin 1
AY237122
7114 bp

Figure 11

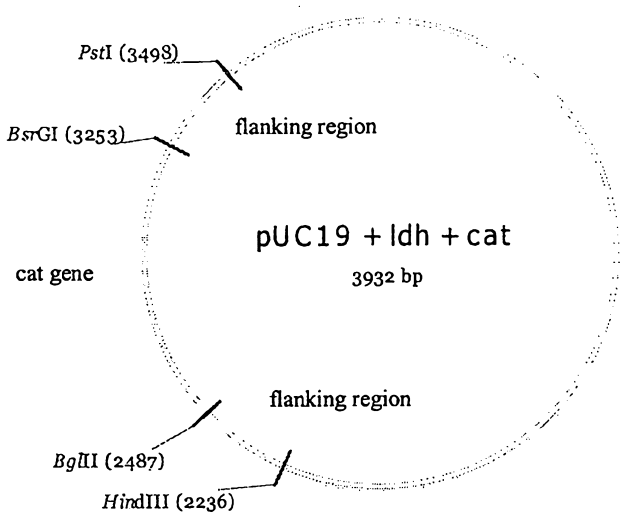
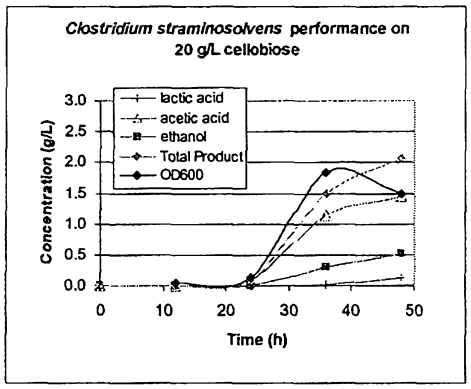

PstI (3498)
BsrGI (3253)
flanking region
pUC19 + ldh + cat
3932 bp
cat gene
flanking region
BglII (2487)
HindIII (2236)

Figure 12A	Figure 12B



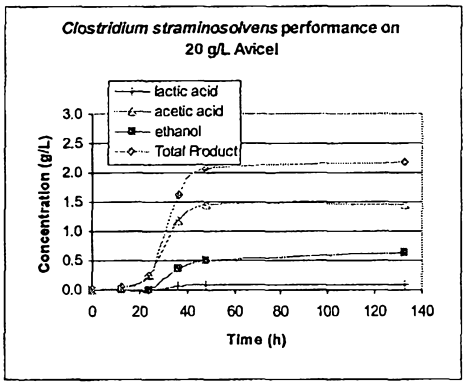
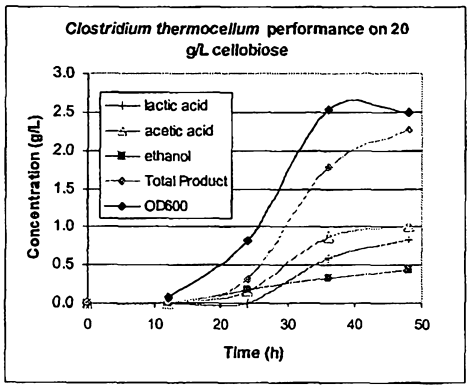


Figure 13A	Figure 13B



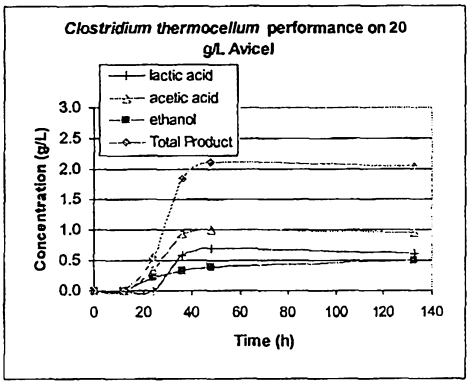
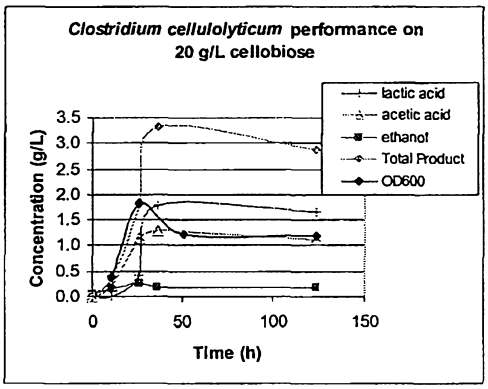


Figure 14A	Figure14B



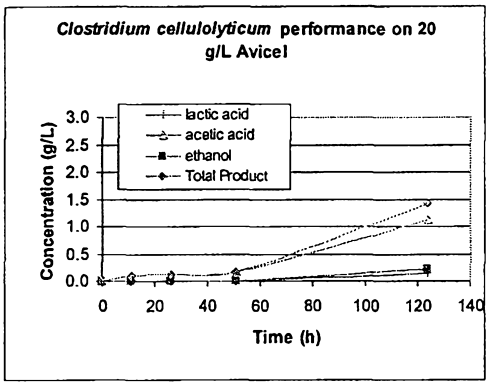
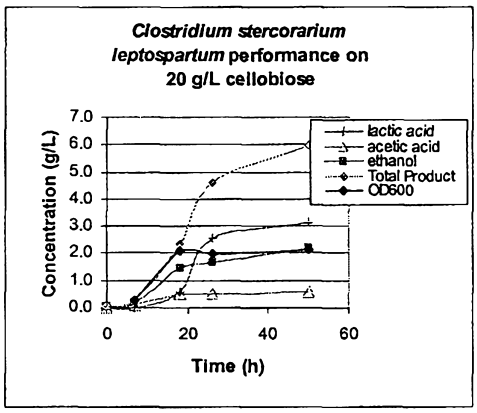


Figure 15A	Figure 15B



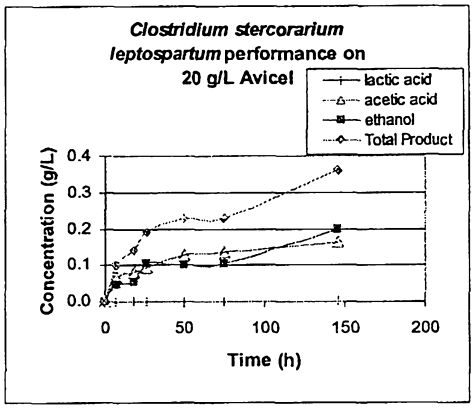
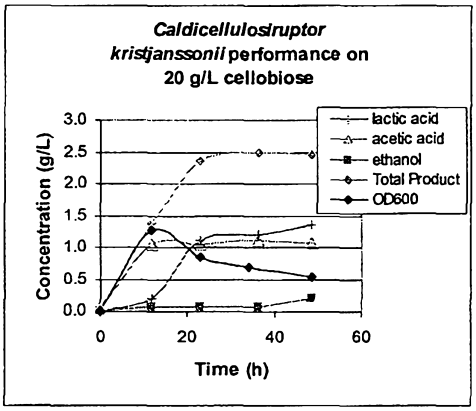


Figure 16A	Figure 16B



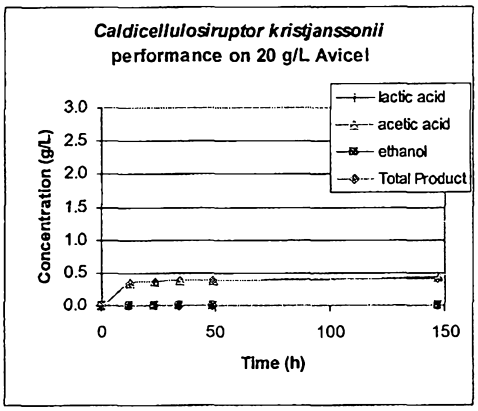
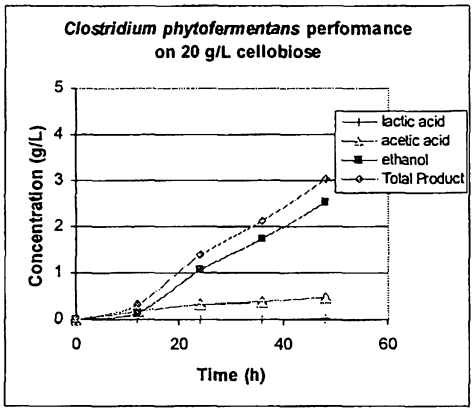


Figure 17A	Figure 17B



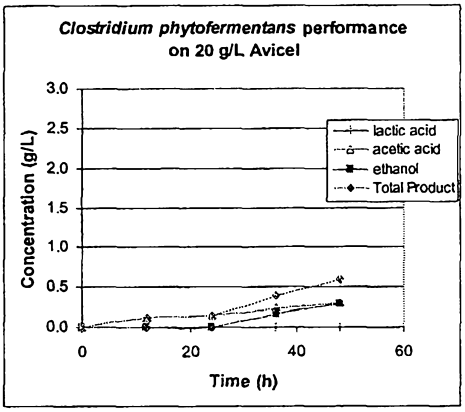


Figure 18

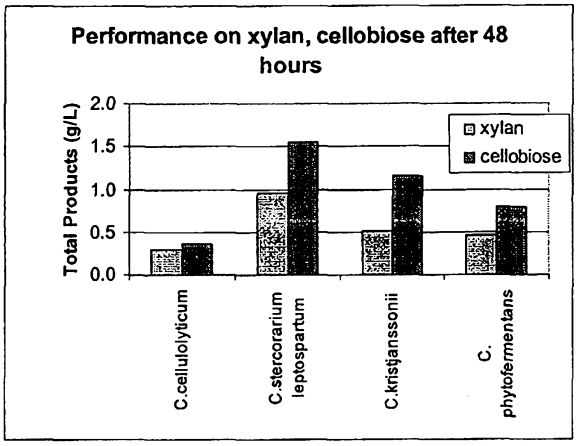

Performance on xylan, cellobiose after 48 hours
Total Products (g/L)
2.0
1.5
1.0
0.5
0.0
xylan
cellobiose
C.cellulolyticum
C.stercorarium leptospartum
C.kristjanssonii
C. phytofermentans

Figure 19

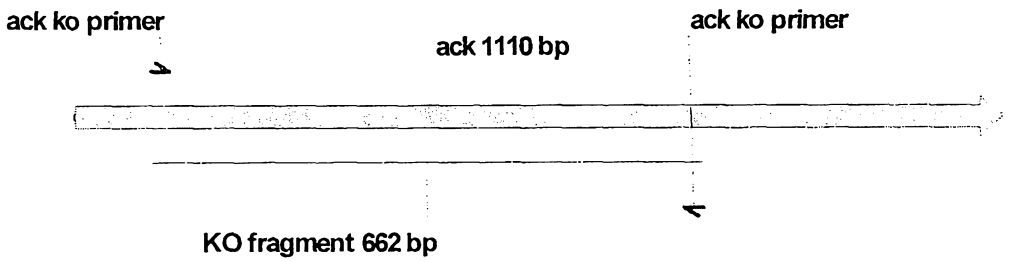
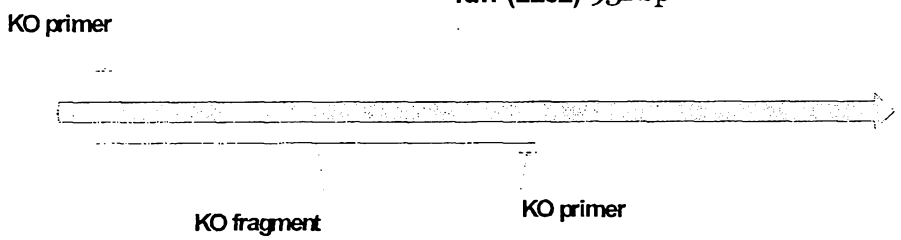
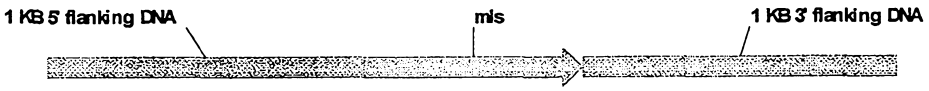


Figure 20



ldh 2262 region for single cross over knockout

Figure 21



C. cellulolyticum ldh 2262 double cross over KO fragment

Figure 22

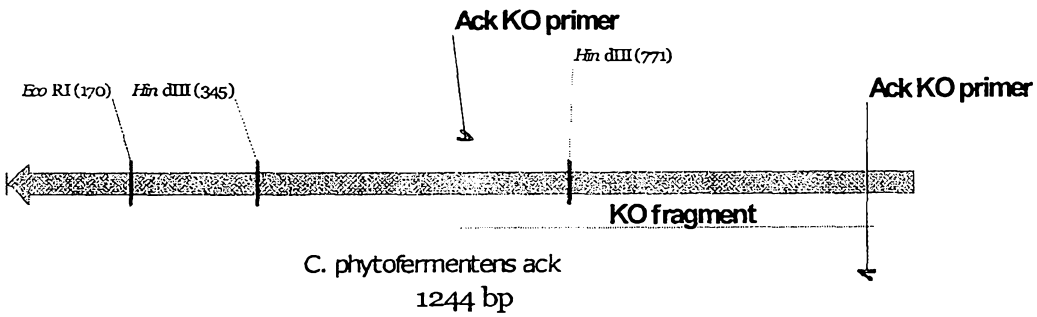
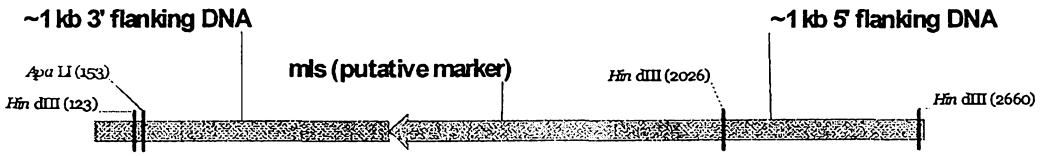
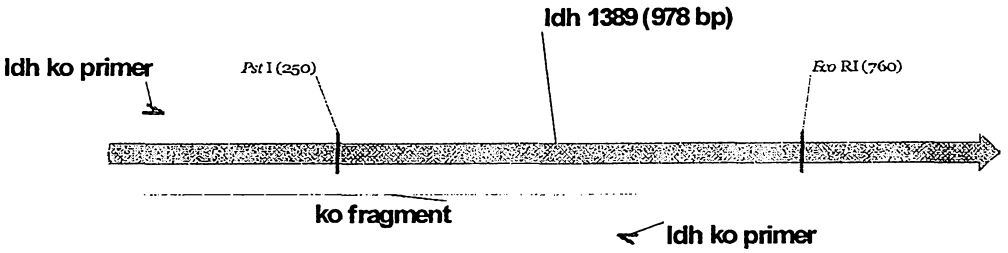

Ack KO primer
Hin dIII (771)
Eco RI (170)
Hin dIII (345)
Ack KO primer
KO fragment
C. phytofermentens ack
1244 bp

Figure 23



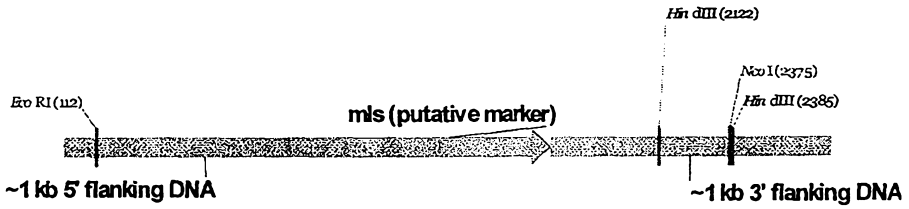
C. phytofermentens ack double cross over KO fragment
2676 bp

Figure 24



C. phytofermentens ldh 1389 single cross over fragment

Figure 25



C. phytpfermentens double crossover ko fragment

2732 bp

Figure 26

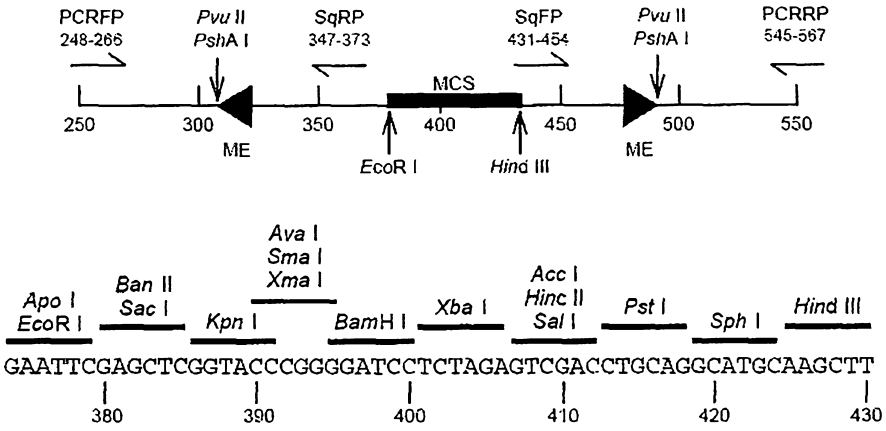

PCRFP
248-266
Pvu II
PshA I
SqRP
347-373
SqFP
431-454
Pvu II
PshA I
PCRRP
545-567
MCS
250
300
350
400
450
500
550
ME
ME
EcoR I
Hind III
Ava I
Sma I
Xma I
Apo I
EcoR I
Ban II
Sac I
Kpn I
BamH I
Xba I
Acc I
Hinc II
Sal I
Pst I
Sph I
Hind III
GAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTT
380
390
400
410
420
430

Figure 27

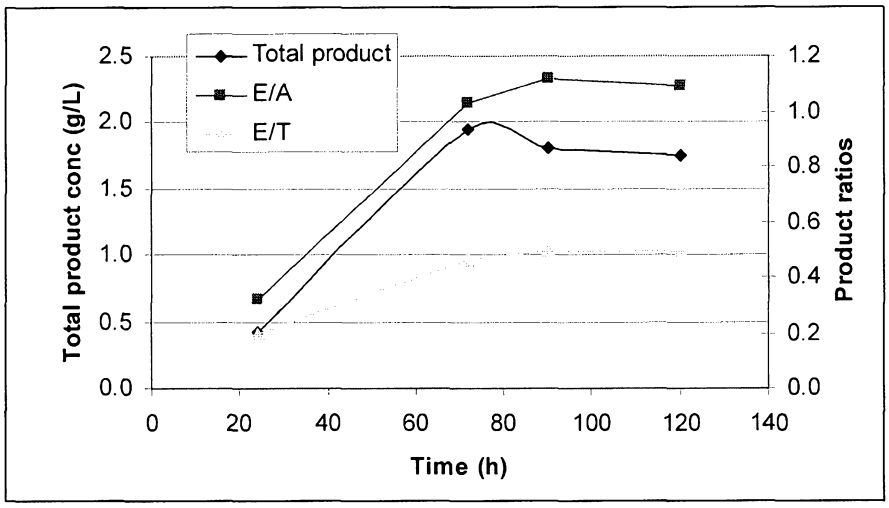

Total product
E/A
E/T
Total product conc (g/L)
Product ratios
Time (h)
2.5
2.0
1.5
1.0
0.5
0.0
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
140

Figure 28

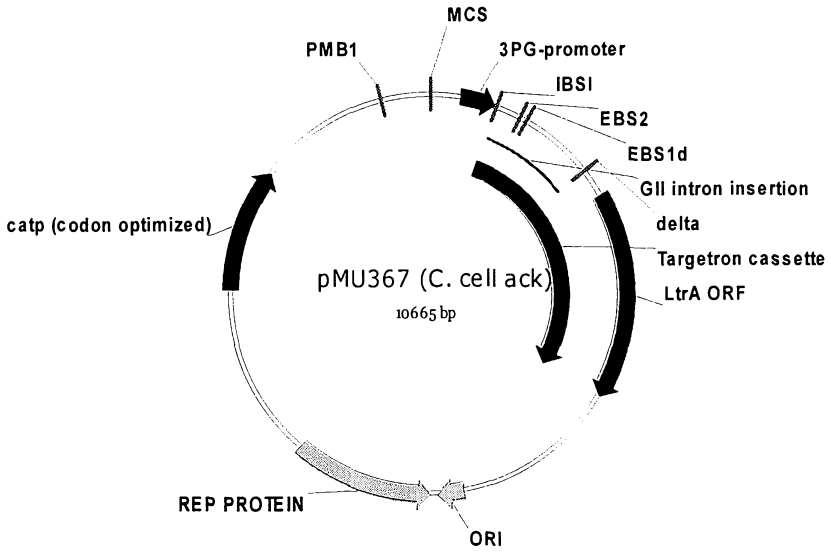

MCS
PMB1
3PG-promoter
IBSI
EBS2
EBS1d
GII intron insertion
delta
Targetron cassette
LtrA ORF
catp (codon optimized)
pMU367 (C. cell ack)
10665 bp
REP PROTEIN
ORI

Figure 29

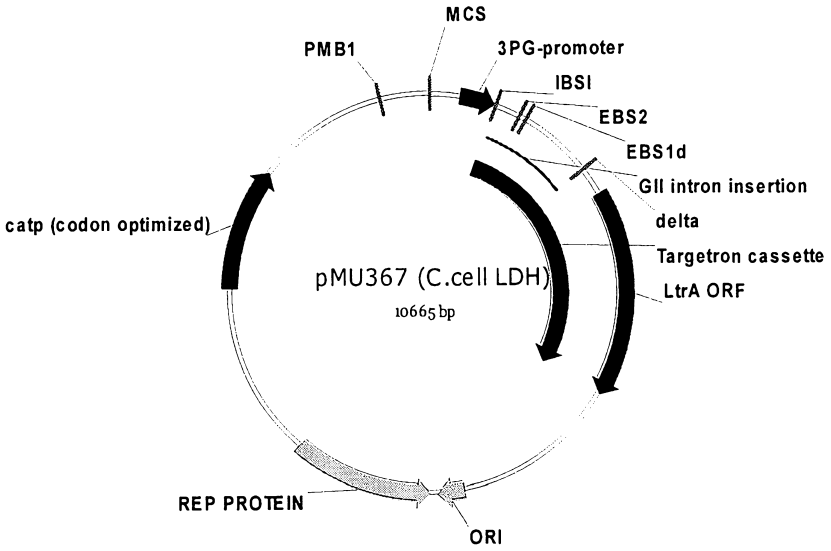
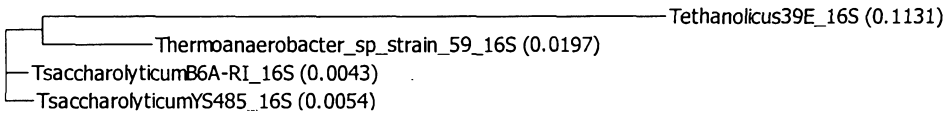
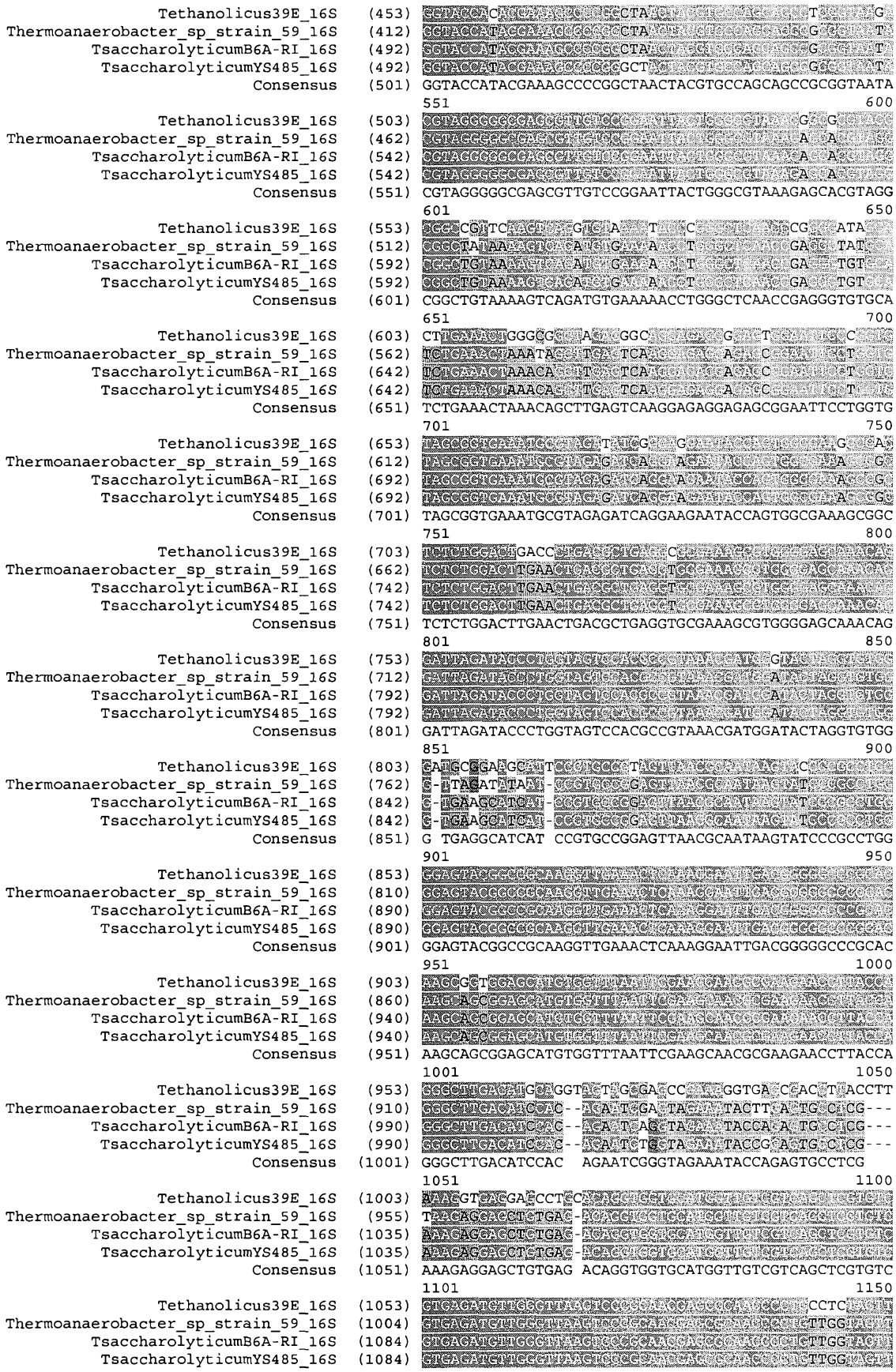

MCS
PMB1
3PG-promoter
IBSI
EBS2
EBS1d
GII intron insertion
delta
Targetron cassette
LtrA ORF
catp (codon optimized)
pMU367 (C.cell LDH)
10665 bp
REP PROTEIN
ORI

Figure 30

```
Percent identity of 16S sequence to T. saccharolyticum JW/SL-YS485:
Thermoanaerobacterium saccharolyticum B6A-RI - 99.0%
Thermoanaerobacter sp. strain 59  - 95.7%
Thermoanaerobacter pseudoethanolicus 39E - 83.7%
```




```
                                      1                                                  50
                Tethanolicus39E_16S     (1) ------CCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAG
Thermoanaerobacter_sp_strain_59_16S     (1) --------------------------------------------------
         TsaccharolyticumB6A-RI_16S     (1) TTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAG
          TsaccharolyticumYS485_16S     (1) TTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAG
                          Consensus     (1) TTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAG
                                     51                                                 100
                Tethanolicus39E_16S    (45) TCGAGCGGTCCGGC----------------------------AGCCAA-
Thermoanaerobacter_sp_strain_59_16S     (1) ----------------------------CTACACATGCAGTCGAGCGAA-
         TsaccharolyticumB6A-RI_16S    (51) TCGAGCGATCCGGNACTCAATTAAGGGCTTACAGAAAAAGA-GAGCAGAAA
          TsaccharolyticumYS485_16S    (51) TCGAGCGATCCGGCACTCAACTAAGCGCTTACAGAAAAAGA-GAGCAGAAA
                          Consensus    (51) TCGAGCGATCCGGCACTCAA TAAGCGCTTACAGAAAAAGA GAGCGAAA
                                    101                                                 150
                Tethanolicus39E_16S    (65) ----------CTTANCNCNGGAGCGGATAGCGGCGAGCGGTGAGTAAC
Thermoanaerobacter_sp_strain_59_16S    (22) GGGAGTA----CTACCGTACCAACT---TAGCGGCGGACGGGTGAGTAAC
         TsaccharolyticumB6A-RI_16S   (100) NTGAGTAAACGCAAGTTGACTGCCGGATAGCGGCGGACGGGTGAGTAAC
          TsaccharolyticumYS485_16S   (100) ATGAGTAAACGCAAGTTGACTGCCGGATAGCGGCGGACGGGTGAGTAAC
                          Consensus   (101)  TGAGTAAACGCTAAGTTGAGTGCCGGATAGCGGCGGACGGGTGAGTAAC
                                    151                                                 200
                Tethanolicus39E_16S   (105) GCGTGGCGCAACCTACCCTTAGACCGGGATAACACCTCGAAAGGGGTGCT
Thermoanaerobacter_sp_strain_59_16S    (65) GCGTGGAGAATTTTCGGATGTAGACCGGGATAACACCTCGAAAGGGGTGCT
         TsaccharolyticumB6A-RI_16S   (150) GCGTGGCACAATCTACCCTGTAGTTTGGGATAACACCTCGAAAGGGGTGCT
          TsaccharolyticumYS485_16S   (150) GCGTGGACAATCTACCCTGTAGTTTGGGATAACACCTCGAAAGGGGTGCT
                          Consensus   (151) GCGTGGACAATCTACCCTGTAGTTTGGGATAACACCTCGAAAGGGGTGCT
                                    201                                                 250
                Tethanolicus39E_16S   (155) AATACTGGATAAGCTGCCTTGTAGGGCATGGTAAGAGCAGCCGCGTACCGG
Thermoanaerobacter_sp_strain_59_16S   (115) AATACCGGATAATGTCGAGAAGCGGCATCGCTTTTTGAAGAAAGGACAG-
         TsaccharolyticumB6A-RI_16S   (200) AATACCGGATAATGTCAAGAAGTGGCATCACTTTTTGAAGAAAGGACA--
          TsaccharolyticumYS485_16S   (200) AATACCGGATAATGTCAAGAAGTGGCATCACTTTTTGAAGAAAGGACA--
                          Consensus   (201) AATACCGGATAATGTCAAGAAGTGGCATCGCTTTTTGAAGAAAGGAGAG
                                    251                                                 300
                Tethanolicus39E_16S   (205) GACTAGCGGTTAAGGATGAGCCAGGCGTGCGATCAGCTAGTTGGTA-GGGT
Thermoanaerobacter_sp_strain_59_16S   (164) -AAT-GGGCTATAGGATGAGTCCGCGTCCCATTAGCTAGTTGGCGA-GGGT
         TsaccharolyticumB6A-RI_16S   (248) -AAT-GGGCTATAGGATGAGTCCGCGTCCCATTAGCTAGTTGGCG-GGGT
          TsaccharolyticumYS485_16S   (248) -AAT-GGGCTATAGGATGAGTCCGCGTCCCATTAGCTAGTTGGCG-GGGT
                          Consensus   (251)  AAT CCGCTATAGGATGAGTCCGCGTCCCATTAGCTAGTTGGCG GGGT
                                    301                                                 350
                Tethanolicus39E_16S   (254) AACGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGTCGGC
Thermoanaerobacter_sp_strain_59_16S   (212) AAAAGGCCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGAACGGC
         TsaccharolyticumB6A-RI_16S   (295) AAAAGGCCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGAACGGC
          TsaccharolyticumYS485_16S   (295) AAAAGGCCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGAACGGC
                          Consensus   (301) AAAAGCCCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGAACGGC
                                    351                                                 400
                Tethanolicus39E_16S   (304) CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGG
Thermoanaerobacter_sp_strain_59_16S   (262) CACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
         TsaccharolyticumB6A-RI_16S   (345) CACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
          TsaccharolyticumYS485_16S   (345) CACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
                          Consensus   (351) CACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
                                    401                                                 450
                Tethanolicus39E_16S   (354) GAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCGACGCCGCGTGAGCGA
Thermoanaerobacter_sp_strain_59_16S   (312) GAATATTGTGCAATGGGGGAAACCCTGACACAGCGACGCCGCGTGAGTGA
         TsaccharolyticumB6A-RI_16S   (395) GAATATTGTTCAATGGGGGAAACCCTGACACAGCGACGCCGCGTGAGCGA
          TsaccharolyticumYS485_16S   (395) GAATATTGTGCAATGGGGGAAACCCTGACACAGCGACGCCGCGTGAGCGA
                          Consensus   (401) GAATATTGTGCAATGGGGGAAACCCTGACACAGCGACGCCGCGTGAGCGA
                                    451                                                 500
                Tethanolicus39E_16S   (404) GCAAGGCCTTCGGGTCGTAAAGCTCGATAGTTGTCGAAGAAGGGA-TGAC
Thermoanaerobacter_sp_strain_59_16S   (362) AGAAGGCCTTCGGGTCGTAAAGCTCAATAGTATGGGAAGATAGAACAAATGAC
         TsaccharolyticumB6A-RI_16S   (445) AGAAGGCCTTCGGGTCGTAAAGCTCAATAGTATGGGAAGATAG---TGAC
          TsaccharolyticumYS485_16S   (445) AGAAGGCCTTCGGGTCGTAAAGCTCAATAGTATGGGAAGATAG---TGAC
                          Consensus   (451) AGAAGGCCTTCGGGTCGTAAAGCTCAATAGTATGGGAAGATAG A TGAC
                                    501                                                 550
```


Tethanolicus39E_16S (453)
Thermoanaerobacter_sp_strain_59_16S (412)
TsaccharolyticumB6A-RI_16S (492)
TsaccharolyticumYS485_16S (492)
Consensus (501) GGTACCATACGAAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATA
551 600
Tethanolicus39E_16S (503)
Thermoanaerobacter_sp_strain_59_16S (462)
TsaccharolyticumB6A-RI_16S (542)
TsaccharolyticumYS485_16S (542)
Consensus (551) CGTAGGGGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCACGTAGG
601 650
Tethanolicus39E_16S (553)
Thermoanaerobacter_sp_strain_59_16S (512)
TsaccharolyticumB6A-RI_16S (592)
TsaccharolyticumYS485_16S (592)
Consensus (601) CGGCTGTAAAAGTCAGATGTGAAAAACCTGGGCTCAACCGAGGGTGTGCA
651 700
Tethanolicus39E_16S (603)
Thermoanaerobacter_sp_strain_59_16S (562)
TsaccharolyticumB6A-RI_16S (642)
TsaccharolyticumYS485_16S (642)
Consensus (651) TCTGAAACTAAACAGCTTGAGTCAAGGAGAGGAGAGCGGAATTCCTGGTG
701 750
Tethanolicus39E_16S (653)
Thermoanaerobacter_sp_strain_59_16S (612)
TsaccharolyticumB6A-RI_16S (692)
TsaccharolyticumYS485_16S (692)
Consensus (701) TAGCGGTGAAATGCGTAGAGATCAGGAAGAATACCAGTGGCGAAAGCGGC
751 800
Tethanolicus39E_16S (703)
Thermoanaerobacter_sp_strain_59_16S (662)
TsaccharolyticumB6A-RI_16S (742)
TsaccharolyticumYS485_16S (742)
Consensus (751) TCTCTGGACTTGAACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAG
801 850
Tethanolicus39E_16S (753)
Thermoanaerobacter_sp_strain_59_16S (712)
TsaccharolyticumB6A-RI_16S (792)
TsaccharolyticumYS485_16S (792)
Consensus (801) GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTGG
851 900
Tethanolicus39E_16S (803)
Thermoanaerobacter_sp_strain_59_16S (762)
TsaccharolyticumB6A-RI_16S (842)
TsaccharolyticumYS485_16S (842)
Consensus (851) G TGAGGCATCAT CCGTGCCGGAGTTAACGCAATAAGTATCCCGCCTGG
901 950
Tethanolicus39E_16S (853)
Thermoanaerobacter_sp_strain_59_16S (810)
TsaccharolyticumB6A-RI_16S (890)
TsaccharolyticumYS485_16S (890)
Consensus (901) GGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCAC
951 1000
Tethanolicus39E_16S (903)
Thermoanaerobacter_sp_strain_59_16S (860)
TsaccharolyticumB6A-RI_16S (940)
TsaccharolyticumYS485_16S (940)
Consensus (951) AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA
1001 1050
Tethanolicus39E_16S (953)
Thermoanaerobacter_sp_strain_59_16S (910)
TsaccharolyticumB6A-RI_16S (990)
TsaccharolyticumYS485_16S (990)
Consensus (1001) GGGCTTGACATCCAC AGAATCGGGTAGAAATACCAGAGTGCCTCG
1051 1100
Tethanolicus39E_16S (1003)
Thermoanaerobacter_sp_strain_59_16S (955)
TsaccharolyticumB6A-RI_16S (1035)
TsaccharolyticumYS485_16S (1035)
Consensus (1051) AAAGAGGAGCTGTGAG ACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTC
1101 1150
Tethanolicus39E_16S (1053)
Thermoanaerobacter_sp_strain_59_16S (1004)
TsaccharolyticumB6A-RI_16S (1084)
TsaccharolyticumYS485_16S (1084)

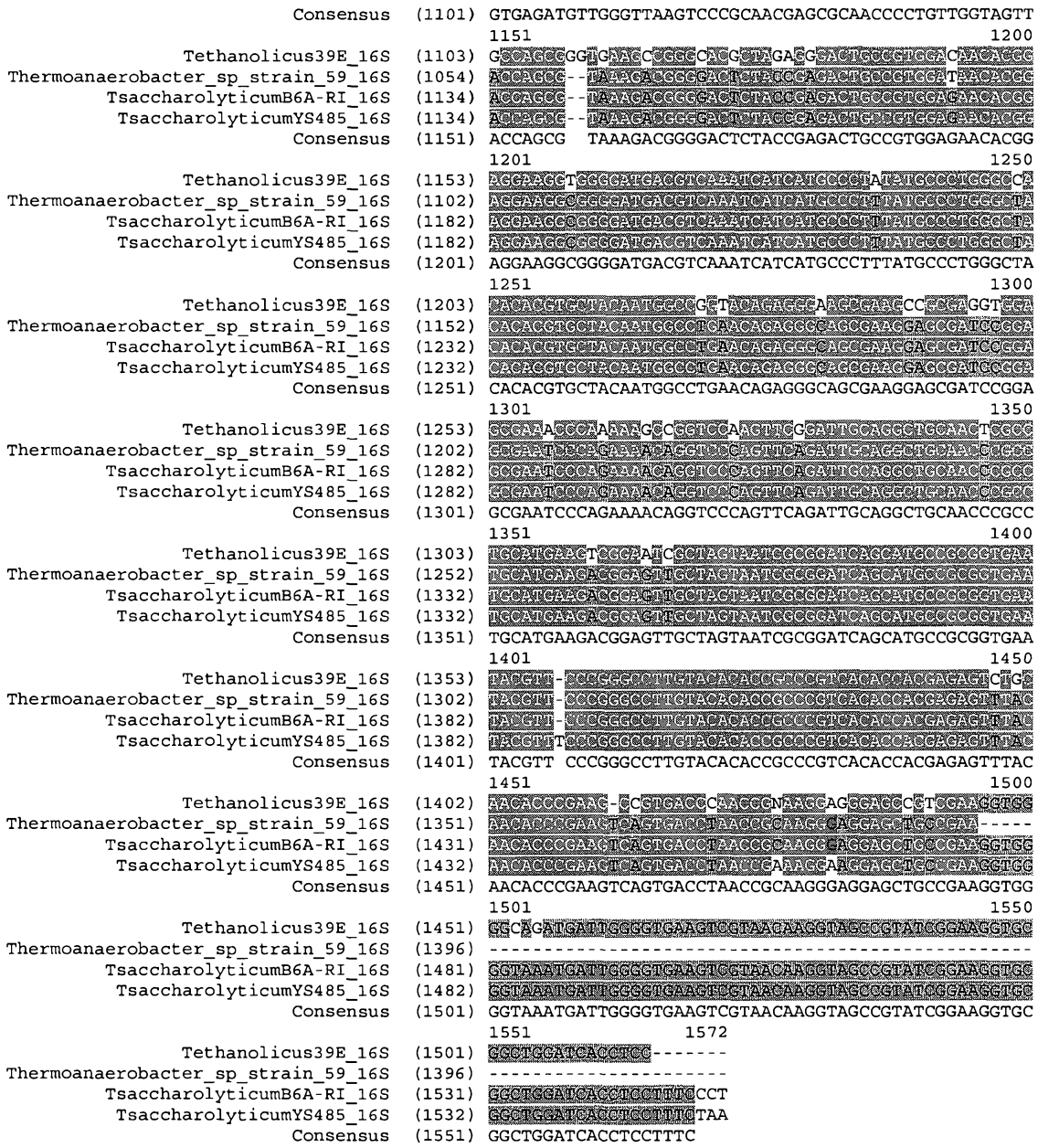
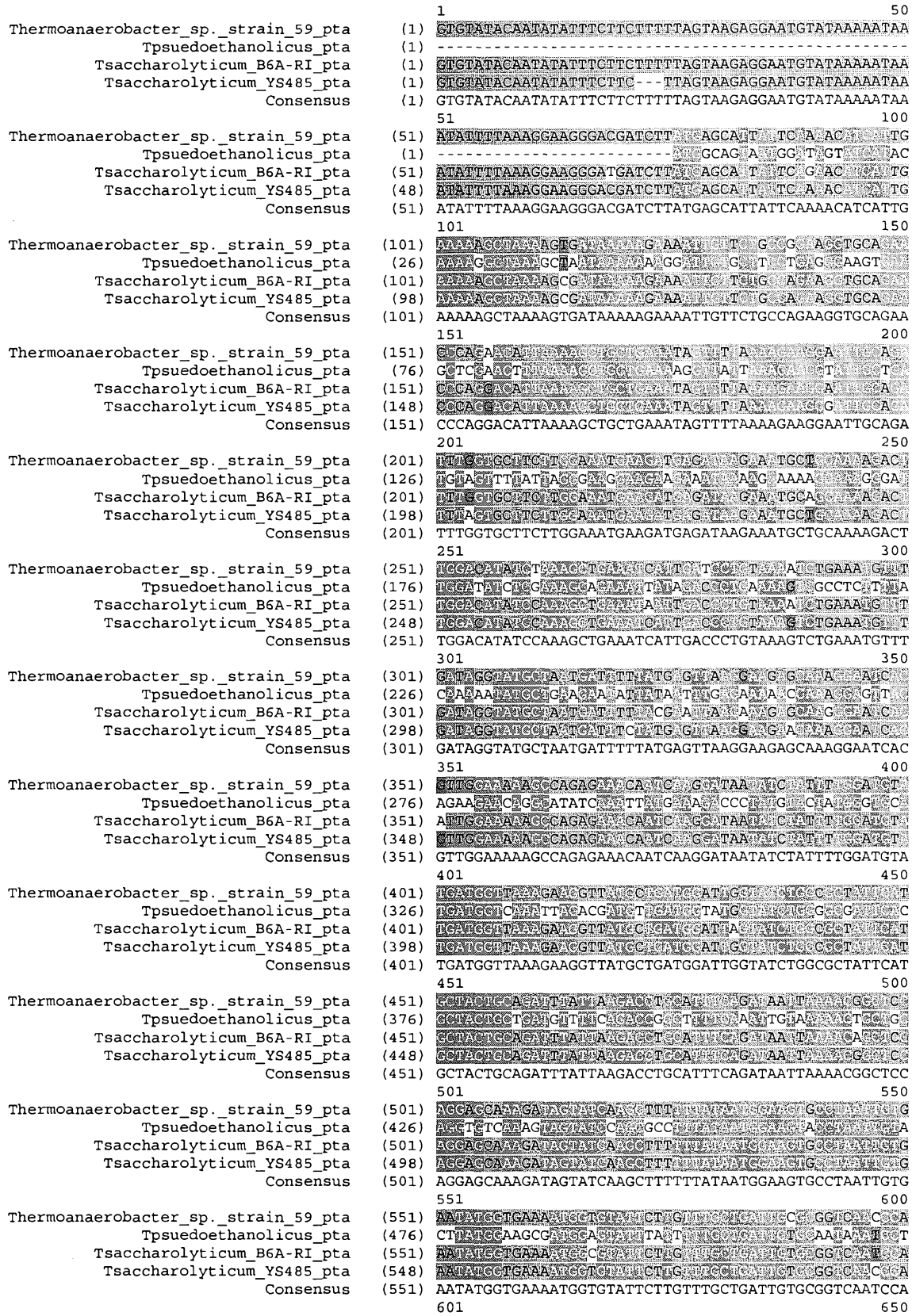
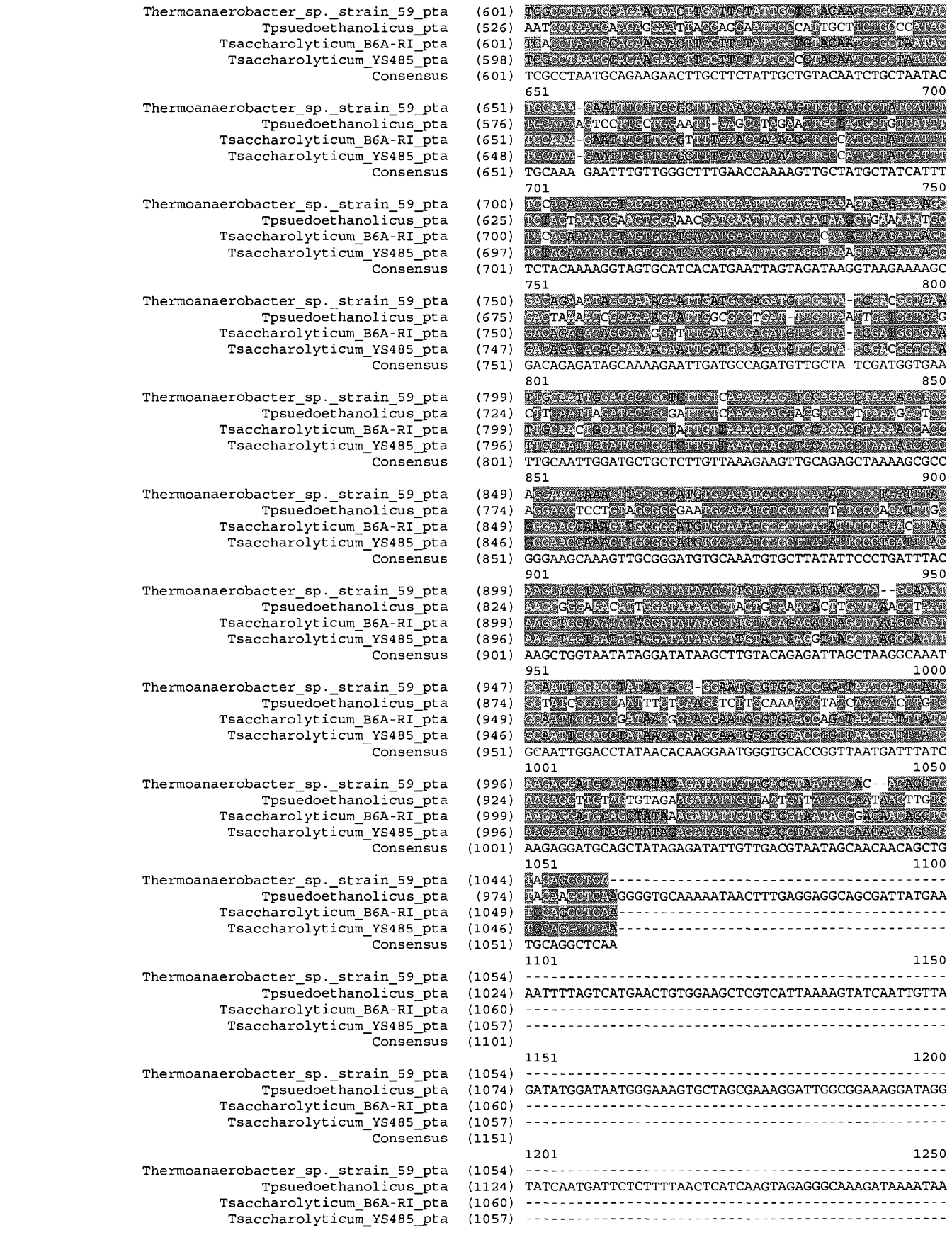

Consensus (1101) GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTTGGTAGTT
1151 1200
Tethanolicus39E_16S (1103)
Thermoanaerobacter_sp_strain_59_16S (1054)
TsaccharolyticumB6A-RI_16S (1134)
TsaccharolyticumYS485_16S (1134)
Consensus (1151) ACCAGCG TAAAGACGGGGACTCTACCGAGACTGCCGTGGAGAACACGG
1201 1250
Tethanolicus39E_16S (1153)
Thermoanaerobacter_sp_strain_59_16S (1102)
TsaccharolyticumB6A-RI_16S (1182)
TsaccharolyticumYS485_16S (1182)
Consensus (1201) AGGAAGGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA
1251 1300
Tethanolicus39E_16S (1203)
Thermoanaerobacter_sp_strain_59_16S (1152)
TsaccharolyticumB6A-RI_16S (1232)
TsaccharolyticumYS485_16S (1232)
Consensus (1251) CACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGA
1301 1350
Tethanolicus39E_16S (1253)
Thermoanaerobacter_sp_strain_59_16S (1202)
TsaccharolyticumB6A-RI_16S (1282)
TsaccharolyticumYS485_16S (1282)
Consensus (1301) GCGAATCCCAGAAAACAGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCC
1351 1400
Tethanolicus39E_16S (1303)
Thermoanaerobacter_sp_strain_59_16S (1252)
TsaccharolyticumB6A-RI_16S (1332)
TsaccharolyticumYS485_16S (1332)
Consensus (1351) TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA
1401 1450
Tethanolicus39E_16S (1353)
Thermoanaerobacter_sp_strain_59_16S (1302)
TsaccharolyticumB6A-RI_16S (1382)
TsaccharolyticumYS485_16S (1382)
Consensus (1401) TACGTT CCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTAC
1451 1500
Tethanolicus39E_16S (1402)
Thermoanaerobacter_sp_strain_59_16S (1351)
TsaccharolyticumB6A-RI_16S (1431)
TsaccharolyticumYS485_16S (1432)
Consensus (1451) AACACCCGAAGTCAGTGACCTAACCGCAAGGGAGGAGCTGCCGAAGGTGG
1501 1550
Tethanolicus39E_16S (1451)
Thermoanaerobacter_sp_strain_59_16S (1396)
TsaccharolyticumB6A-RI_16S (1481)
TsaccharolyticumYS485_16S (1482)
Consensus (1501) GGTAAATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGC
1551 1572
Tethanolicus39E_16S (1501)
Thermoanaerobacter_sp_strain_59_16S (1396)
TsaccharolyticumB6A-RI_16S (1531)
TsaccharolyticumYS485_16S (1532)
Consensus (1551) GGCTGGATCACCTCCTTTC

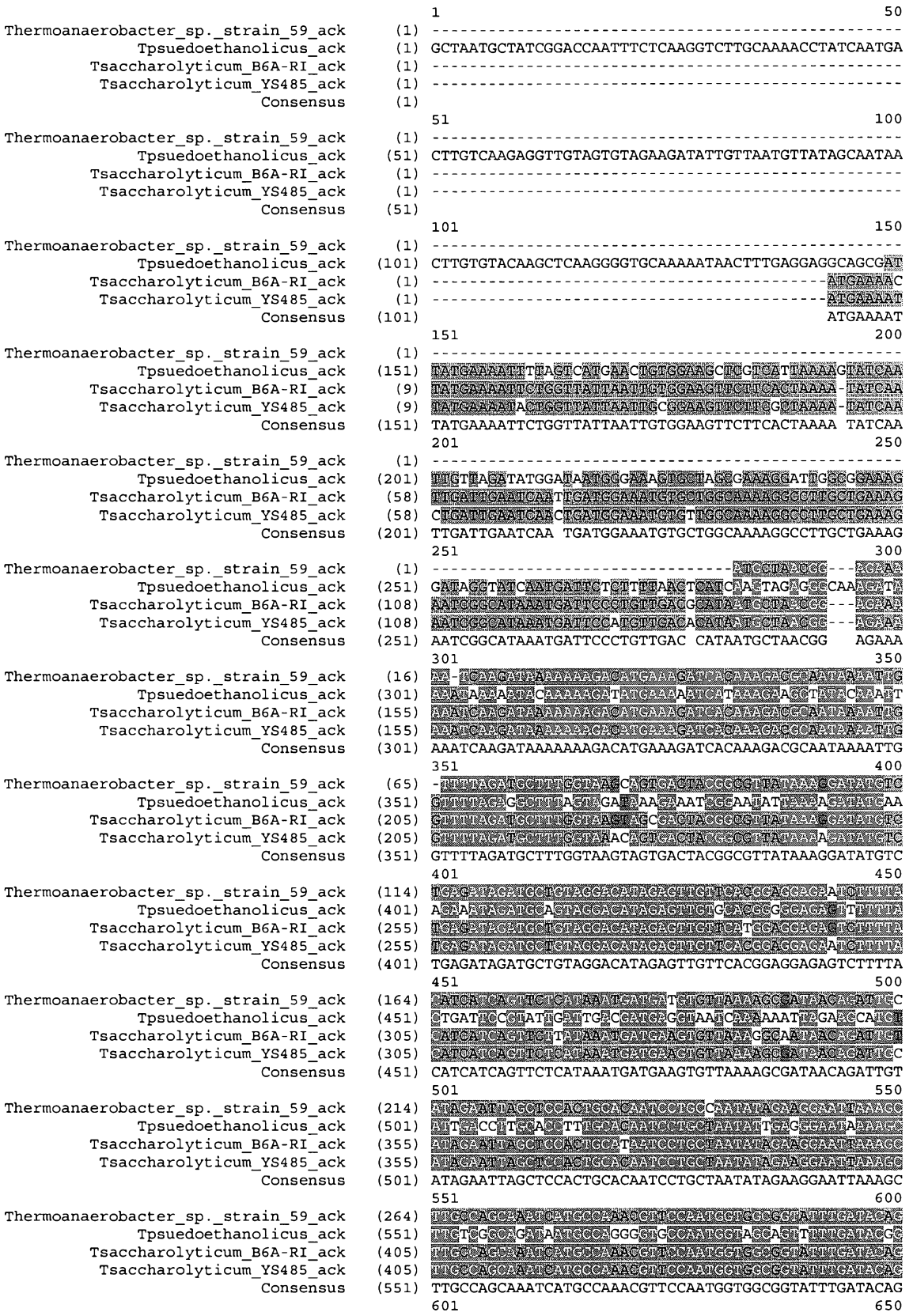
Figure 31

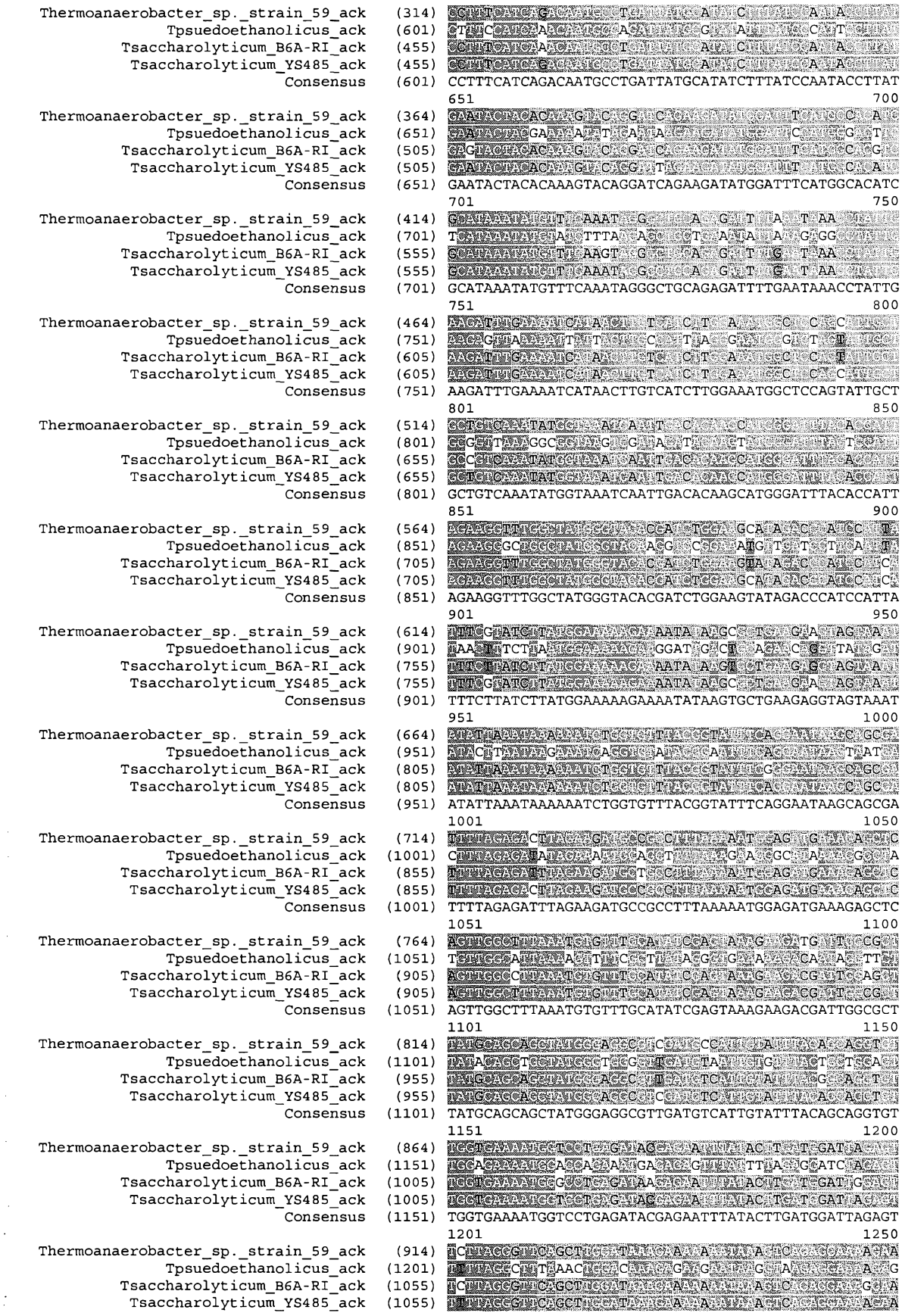

Thermoanaerobacter_sp._strain_59_pta
Tpsuedoethanolicus_pta
Tsaccharolyticum_B6A-RI_pta
Tsaccharolyticum_YS485_pta
Consensus


Thermoanaerobacter_sp._strain_59_pta
Tpsuedoethanolicus_pta
Tsaccharolyticum_B6A-RI_pta
Tsaccharolyticum_YS485_pta
Consensus

```
                                Consensus  (1201)
                                                   1251                1273
Thermoanaerobacter_sp._strain_59_pta  (1054) ----------------------
                    Tpsuedoethanolicus_pta  (1174) AAATACAAAAAGATATGAAAAAT
             Tsaccharolyticum_B6A-RI_pta  (1060) ----------------------
              Tsaccharolyticum_YS485_pta  (1057) ----------------------
                                Consensus  (1251)
```

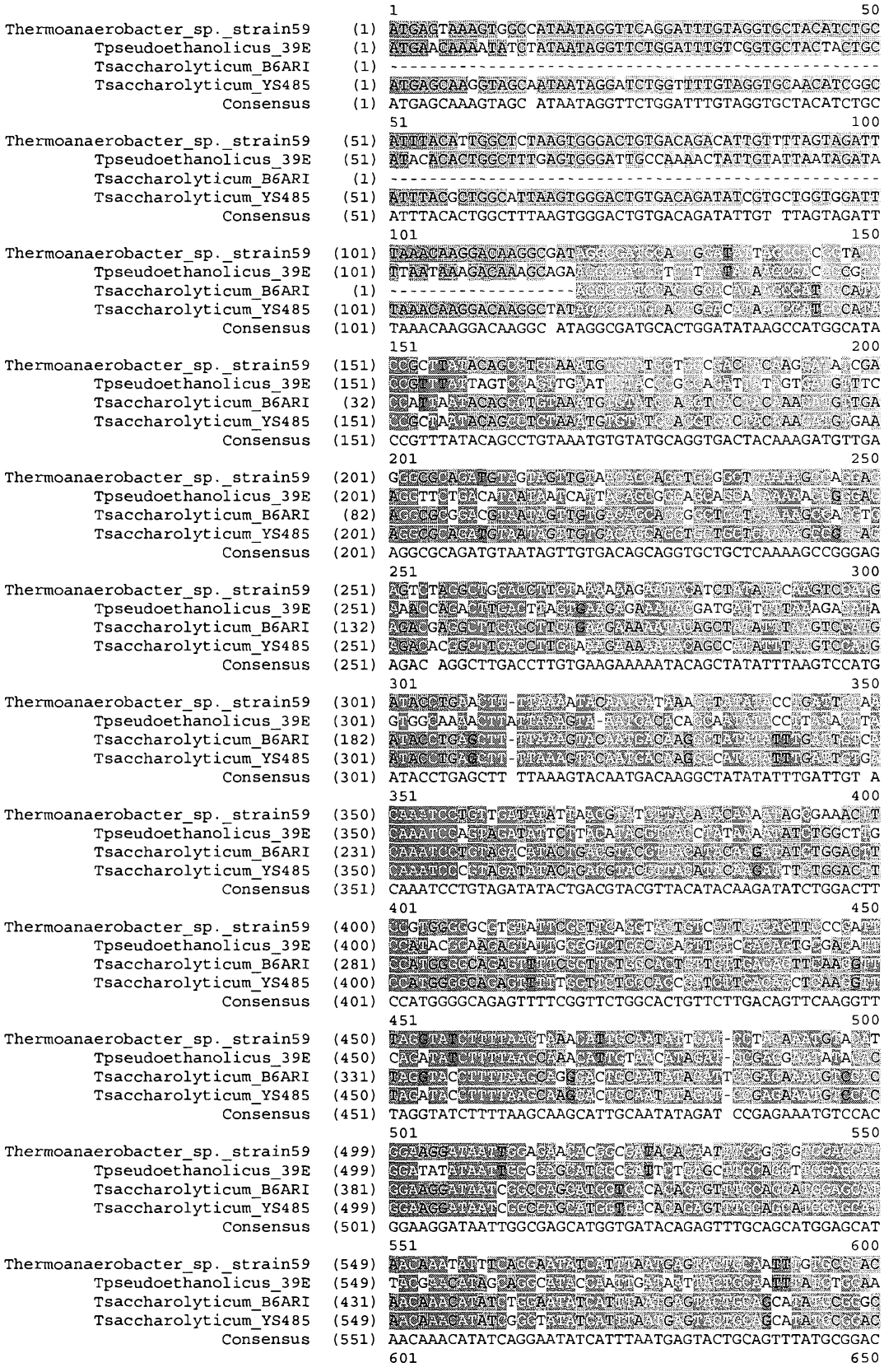
Figure 32


1 50
Thermoanaerobacter_sp._strain_59_ack (1)
Tpsuedoethanolicus_ack (1) GCTAATGCTATCGGACCAATTTCTCAAGGTCTTGCAAAACCTATCAATGA
Tsaccharolyticum_B6A-RI_ack (1)
Tsaccharolyticum_YS485_ack (1)
Consensus (1)
51 100
Thermoanaerobacter_sp._strain_59_ack (1)
Tpsuedoethanolicus_ack (51) CTTGTCAAGAGGTTGTAGTGTAGAAGATATTGTTAATGTTATAGCAATAA
Tsaccharolyticum_B6A-RI_ack (1)
Tsaccharolyticum_YS485_ack (1)
Consensus (51)
101 150
Thermoanaerobacter_sp._strain_59_ack (1)
Tpsuedoethanolicus_ack (101) CTTGTGTACAAGCTCAAGGGGTGCAAAAATAACTTTGAGGAGGCAGCGAT
Tsaccharolyticum_B6A-RI_ack (1) ATGAAAAC
Tsaccharolyticum_YS485_ack (1) ATGAAAAT
Consensus (101) ATGAAAAT
151 200
Thermoanaerobacter_sp._strain_59_ack (1)
Tpsuedoethanolicus_ack (151) TATGAAAATTTTAGTCATGAACTGTGGAAGCTCGTCATTAAAAGTATCAA
Tsaccharolyticum_B6A-RI_ack (9) TATGAAAATTCTGGTTATTAATTGTGGAAGTTCTTCACTAAAA-TATCAA
Tsaccharolyticum_YS485_ack (9) TATGAAAAATACTGGTTATTAATTGCGGAAGTTCTTGCGTAAAA-TATCAA
Consensus (151) TATGAAAATTCTGGTTATTAATTGTGGAAGTTCTTCACTAAAA TATCAA
201 250
Thermoanaerobacter_sp._strain_59_ack (1)
Tpsuedoethanolicus_ack (201) TTGTTAGATATGGATAATGGAAAGTGCTAGGGAAAGGATTGGCGGAAAG
Tsaccharolyticum_B6A-RI_ack (58) TTGATTGAATCAATTGATGCAAATGTGCTGGCAAAAGGCCTTGCTGAAAG
Tsaccharolyticum_YS485_ack (58) CTGATTGAATCAACTGATGGAAATGTGTTGGCAAAAGGCCTTGCTGAAAG
Consensus (201) TTGATTGAATCAA TGATGGAAATGTGCTGGCAAAAGGCCTTGCTGAAAG
251 300
Thermoanaerobacter_sp._strain_59_ack (1) ATGCTAACGG---AGAA
Tpsuedoethanolicus_ack (251) GATAGGTATCAATGATTCTCTTTTAACTCATCAATAGAGCGCAAAGATA
Tsaccharolyticum_B6A-RI_ack (108) AATGGGCATAAATGATTCCGTGTTGACGCATAATGCTAACGG---AGAA
Tsaccharolyticum_YS485_ack (108) AATCGCCATAAATGATTCCATGTTGACACATAATGCTAACGG---AGAA
Consensus (251) AATCGGCATAAATGATTCCCTGTTGAC CATAATGCTAACGG AGAAA
301 350
Thermoanaerobacter_sp._strain_59_ack (16) AA-TGAAGATAAAAAAAGACATGAAAGATCACAAAGACGCAATAAAATTG
Tpsuedoethanolicus_ack (301) AAATAAAAATACATAAAGATATGAAAAATCATAAAGAACGTATACAAATT
Tsaccharolyticum_B6A-RI_ack (155) AAATCAAGATAAAAAAAGACATGAAAGATCACAAAGACGCAATAAAATTG
Tsaccharolyticum_YS485_ack (155) AAATCAAGATAAAAAAAGACATGAAAGATCACAAAGACGCAATAAAATTG
Consensus (301) AAATCAAGATAAAAAAAGACATGAAAGATCACAAAGACGCAATAAAATTG
351 400
Thermoanaerobacter_sp._strain_59_ack (65) -TTTTAGATGCTTTGGTAAGCAGTGACTACGGCGTTATAAAGGATATGTC
Tpsuedoethanolicus_ack (351) GTTTTAGATGCTTTGGTAAGTAGTGACTACGGCGTTATAAAGGATATGAA
Tsaccharolyticum_B6A-RI_ack (205) GTTTTAGATGCTTTGGTAAGTAGCGACTACGGCGTTATAAAGGATATGTC
Tsaccharolyticum_YS485_ack (205) GTTTTAGATGCTTTGGTAAACAGTGACTACGGCGTTATAAAGGATATGTC
Consensus (351) GTTTTAGATGCTTTGGTAAGTAGTGACTACGGCGTTATAAAGGATATGTC
401 450
Thermoanaerobacter_sp._strain_59_ack (114) TGAGATAGATGCTGTAGGACATAGAGTTGTTCACGGAGGAGAGTCTTTTA
Tpsuedoethanolicus_ack (401) AGAATAGATGCAGTAGGACATAGAGTTGTCACGGCGGAGAGTCTTTTA
Tsaccharolyticum_B6A-RI_ack (255) TCAGATAGATGCTGTAGGACATAGAGTTGTTACTCCACGAGAGTCTTTTA
Tsaccharolyticum_YS485_ack (255) TGAGATAGATGCTGTAGGACATAGAGTTGTTCACGGAGGAGAGTCTTTTA
Consensus (401) TGAGATAGATGCTGTAGGACATAGAGTTGTTCACGGAGGAGAGTCTTTTA
451 500
Thermoanaerobacter_sp._strain_59_ack (164) CATCATCAGTTCTCATAAATGATGAAGTGTTAAAAGCGATAACAGATTGC
Tpsuedoethanolicus_ack (451) CTGATTCCGTATTGATTGACGATGCGTAAACAAAAATTAGAGCCATCT
Tsaccharolyticum_B6A-RI_ack (305) CATCATCAGTTGTTATAAATGATGAACGTTAAAGCGAATAACAGATTGT
Tsaccharolyticum_YS485_ack (305) CATCATCAGTTCTCATAAATGATGAAGTGTTAAAAGCGATAACAGATTGC
Consensus (451) CATCATCAGTTCTCATAAATGATGAAGTGTTAAAAGCGATAACAGATTGT
501 550
Thermoanaerobacter_sp._strain_59_ack (214) ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC
Tpsuedoethanolicus_ack (501) ATTGACCTTCGACTTTGCACAATCCTGCTAATATTGAGGGAATAAAAGC
Tsaccharolyticum_B6A-RI_ack (355) ATAGAATTAGCTCCACTGCATAATCCTGCTAATATAGAAGGAATTAAAGC
Tsaccharolyticum_YS485_ack (355) ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC
Consensus (501) ATAGAATTAGCTCCACTGCACAATCCTGCTAATATAGAAGGAATTAAAGC
551 600
Thermoanaerobacter_sp._strain_59_ack (264) TTCGCAGCAAATCATGCCAAACGTTCCAATGGTGGCGGTATTTGATACAG
Tpsuedoethanolicus_ack (551) TTGCTGGCAGATAATGCCAGGGGTGCCAATGGTAGCAGTTTTTGATACGC
Tsaccharolyticum_B6A-RI_ack (405) TTGCCAGCAAATCATGCCAAACGTTCCAATGGTGGCGGTATTTGATACAC
Tsaccharolyticum_YS485_ack (405) TTGCCAGCAAATCATGCCAAACGTTCCAATGGTGGCGGTATTTGATACAC
Consensus (551) TTGCCAGCAAATCATGCCAAACGTTCCAATGGTGGCGGTATTTGATACAG
601 650


Thermoanaerobacter_sp._strain_59_ack (314)
Tpsuedoethanolicus_ack (601)
Tsaccharolyticum_B6A-RI_ack (455)
Tsaccharolyticum_YS485_ack (455)
Consensus (601) CCTTTCATCAGACAATGCCTGATTATGCATATCTTTATCCAATACCTTAT
651 700
Thermoanaerobacter_sp._strain_59_ack (364)
Tpsuedoethanolicus_ack (651)
Tsaccharolyticum_B6A-RI_ack (505)
Tsaccharolyticum_YS485_ack (505)
Consensus (651) GAATACTACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC
701 750
Thermoanaerobacter_sp._strain_59_ack (414)
Tpsuedoethanolicus_ack (701)
Tsaccharolyticum_B6A-RI_ack (555)
Tsaccharolyticum_YS485_ack (555)
Consensus (701) GCATAAATATGTTTCAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG
751 800
Thermoanaerobacter_sp._strain_59_ack (464)
Tpsuedoethanolicus_ack (751)
Tsaccharolyticum_B6A-RI_ack (605)
Tsaccharolyticum_YS485_ack (605)
Consensus (751) AAGATTTGAAAATCATAACTTGTCATCTTGGAAATGGCTCCAGTATTGCT
801 850
Thermoanaerobacter_sp._strain_59_ack (514)
Tpsuedoethanolicus_ack (801)
Tsaccharolyticum_B6A-RI_ack (655)
Tsaccharolyticum_YS485_ack (655)
Consensus (801) GCTGTCAAATATGGTAAATCAATTGACACAAGCATGGGATTTACACCATT
851 900
Thermoanaerobacter_sp._strain_59_ack (564)
Tpsuedoethanolicus_ack (851)
Tsaccharolyticum_B6A-RI_ack (705)
Tsaccharolyticum_YS485_ack (705)
Consensus (851) AGAAGGTTTGGCTATGGGTACACGATCTGGAAGTATAGACCCATCCATTA
901 950
Thermoanaerobacter_sp._strain_59_ack (614)
Tpsuedoethanolicus_ack (901)
Tsaccharolyticum_B6A-RI_ack (755)
Tsaccharolyticum_YS485_ack (755)
Consensus (901) TTTCTTATCTTATGGAAAAAGAAAATATAAGTGCTGAAGAGGTAGTAAAT
951 1000
Thermoanaerobacter_sp._strain_59_ack (664)
Tpsuedoethanolicus_ack (951)
Tsaccharolyticum_B6A-RI_ack (805)
Tsaccharolyticum_YS485_ack (805)
Consensus (951) ATATTAAATAAAAAATCTGGTGTTTACGGTATTTCAGGAATAAGCAGCGA
1001 1050
Thermoanaerobacter_sp._strain_59_ack (714)
Tpsuedoethanolicus_ack (1001)
Tsaccharolyticum_B6A-RI_ack (855)
Tsaccharolyticum_YS485_ack (855)
Consensus (1001) TTTTAGAGATTTAGAAGATGCCGCCTTTAAAAATGGAGATGAAAGAGCTC
1051 1100
Thermoanaerobacter_sp._strain_59_ack (764)
Tpsuedoethanolicus_ack (1051)
Tsaccharolyticum_B6A-RI_ack (905)
Tsaccharolyticum_YS485_ack (905)
Consensus (1051) AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGACGATTGGCGCT
1101 1150
Thermoanaerobacter_sp._strain_59_ack (814)
Tpsuedoethanolicus_ack (1101)
Tsaccharolyticum_B6A-RI_ack (955)
Tsaccharolyticum_YS485_ack (955)
Consensus (1101) TATGCAGCAGCTATGGGAGGCGTTGATGTCATTGTATTTACAGCAGGTGT
1151 1200
Thermoanaerobacter_sp._strain_59_ack (864)
Tpsuedoethanolicus_ack (1151)
Tsaccharolyticum_B6A-RI_ack (1005)
Tsaccharolyticum_YS485_ack (1005)
Consensus (1151) TGGTGAAAATGGTCCTGAGATACGAGAATTTATACTTGATGGATTAGAGT
1201 1250
Thermoanaerobacter_sp._strain_59_ack (914)
Tpsuedoethanolicus_ack (1201)
Tsaccharolyticum_B6A-RI_ack (1055)
Tsaccharolyticum_YS485_ack (1055)


Consensus (1201) TTTTAGGGTTCAGCTTGGATAAAGAAAAAAATAAAGTCAGAGGAAAAGAA
1251 1300
Thermoanaerobacter_sp._strain_59_ack (964) ACTATTATATCTACGCCGAATTCAAAAGTTAGCGTGATGGTTGTGCCAC
Tpsuedoethanolicus_ack (1251) GAAATTATATCTTCAGAAGATTCAAAAGTTTAAGTTATGCTTAATTCCTAC
Tsaccharolyticum_B6A-RI_ack (1105) ACTATTATATCTACGCCGAATTCAAAAATTAGCGTGATGCTTGTGCCGAC
Tsaccharolyticum_YS485_ack (1105) ACTATTATATCTACGCCGAATTCAAAAGTTAGCGTGATGCTTGTGCCTAC
Consensus (1251) ACTATTATATCTACGCCGAATTCAAAAGTTAGCGTGATGGTTGTGCCTAC
1301 1350
Thermoanaerobacter_sp._strain_59_ack (1014) TAATGAAGAATATCATGATTGCTAAAGATACTGAAAAGATTGTAAAGACTA
Tpsuedoethanolicus_ack (1301) AATGAAGAATATATGATTGCTAAAGATACTGAAATATGCTAAAAGGTT
Tsaccharolyticum_B6A-RI_ack (1155) TAATGAAGAATATATGATTGCTAAAGATACTGAAAAGATTGTAAAGACTA
Tsaccharolyticum_YS485_ack (1155) TAATGAAGAATACATGATTGCTAAAGATACTGAAAAGATTGTAAAGAGTA
Consensus (1301) TAATGAAGAATATATGATTGCTAAAGATACTGAAAAGATTGTAAAGAGTA
1351
Thermoanaerobacter_sp._strain_59_ack (1064) TAAAA---
Tpsuedoethanolicus_ack (1351) TAAAGTAG
Tsaccharolyticum_B6A-RI_ack (1205) TAAAA---
Tsaccharolyticum_YS485_ack (1205) TAAAA---
Consensus (1351) TAAAA

Figure 33


1 50
Thermoanaerobacter_sp._strain59 (1) ATGAGTAAAGTGGCCATAATAGGTTCAGGATTTGTAGGTGCTACATCTGC
Tpseudoethanolicus_39E (1) ATGAACAAAATATCTATAATAGGTTCTGGATTTGTCGGTGCTACTACTGC
Tsaccharolyticum_B6ARI (1) ----------
Tsaccharolyticum_YS485 (1) ATGAGCAAGGTAGCAATAATAGGATCTGGTTTTGTAGGTGCAACATCGGC
Consensus (1) ATGAGCAAAGTAGC ATAATAGGTTCTGGATTTGTAGGTGCTACATCTGC
51 100
Thermoanaerobacter_sp._strain59 (51) ATTTACATTGGCTCTAAGTGGGACTGTGACAGACATTGTTTTAGTAGATT
Tpseudoethanolicus_39E (51) ATACACACTGGCTTTGAGTGGGATTGCCAAAACTATTGTATTAATAGATA
Tsaccharolyticum_B6ARI (1) ----------
Tsaccharolyticum_YS485 (51) ATTTACGCTGGCATTAAGTGGGACTGTGACAGATATCGTGCTGGTGGATT
Consensus (51) ATTTACACTGGCTTTAAGTGGGACTGTGACAGATATTGT TTAGTAGATT
101 150
Thermoanaerobacter_sp._strain59 (101) TAAACAAGGACAAGGCGAT
Tpseudoethanolicus_39E (101) TTAATAAAGACAAAGCAGA
Tsaccharolyticum_B6ARI (1) ----------
Tsaccharolyticum_YS485 (101) TAAACAAGGACAAGGCTAT
Consensus (101) TAAACAAGGACAAGGC ATAGGCGATGCACTGGATATAAGCCATGGCATA
151 200
Thermoanaerobacter_sp._strain59 (151)
Tpseudoethanolicus_39E (151)
Tsaccharolyticum_B6ARI (32)
Tsaccharolyticum_YS485 (151)
Consensus (151) CCGTTTATACAGCCTGTAAATGTGTATGCAGGTGACTACAAAGATGTTGA
201 250
Thermoanaerobacter_sp._strain59 (201)
Tpseudoethanolicus_39E (201)
Tsaccharolyticum_B6ARI (82)
Tsaccharolyticum_YS485 (201)
Consensus (201) AGGCGCAGATGTAATAGTTGTGACAGCAGGTGCTGCTCAAAAGCCGGGAG
251 300
Thermoanaerobacter_sp._strain59 (251)
Tpseudoethanolicus_39E (251)
Tsaccharolyticum_B6ARI (132)
Tsaccharolyticum_YS485 (251)
Consensus (251) AGAC AGGCTTGACCTTGTGAAGAAAAATACAGCTATATTTAAGTCCATG
301 350
Thermoanaerobacter_sp._strain59 (301)
Tpseudoethanolicus_39E (301)
Tsaccharolyticum_B6ARI (182)
Tsaccharolyticum_YS485 (301)
Consensus (301) ATACCTGAGCTT TTAAAGTACAATGACAAGGCTATATATTTGATTGT A
351 400
Thermoanaerobacter_sp._strain59 (350)
Tpseudoethanolicus_39E (350)
Tsaccharolyticum_B6ARI (231)
Tsaccharolyticum_YS485 (350)
Consensus (351) CAAATCCTGTAGATATACTGACGTACGTTACATACAAGATATCTGGACTT
401 450
Thermoanaerobacter_sp._strain59 (400)
Tpseudoethanolicus_39E (400)
Tsaccharolyticum_B6ARI (281)
Tsaccharolyticum_YS485 (400)
Consensus (401) CCATGGGGCAGAGTTTTCGGTTCTGGCACTGTTCTTGACAGTTCAAGGTT
451 500
Thermoanaerobacter_sp._strain59 (450)
Tpseudoethanolicus_39E (450)
Tsaccharolyticum_B6ARI (331)
Tsaccharolyticum_YS485 (450)
Consensus (451) TAGGTATCTTTTAAGCAAGCATTGCAATATAGAT CCGAGAAATGTCCAC
501 550
Thermoanaerobacter_sp._strain59 (499)
Tpseudoethanolicus_39E (499)
Tsaccharolyticum_B6ARI (381)
Tsaccharolyticum_YS485 (499)
Consensus (501) GGAAGGATAATTGGCGAGCATGGTGATACAGAGTTTGCAGCATGGAGCAT
551 600
Thermoanaerobacter_sp._strain59 (549)
Tpseudoethanolicus_39E (549)
Tsaccharolyticum_B6ARI (431)
Tsaccharolyticum_YS485 (549)
Consensus (551) AACAAACATATCAGGAATATCATTTAATGAGTACTGCAGTTTATGCGGAC
601 650

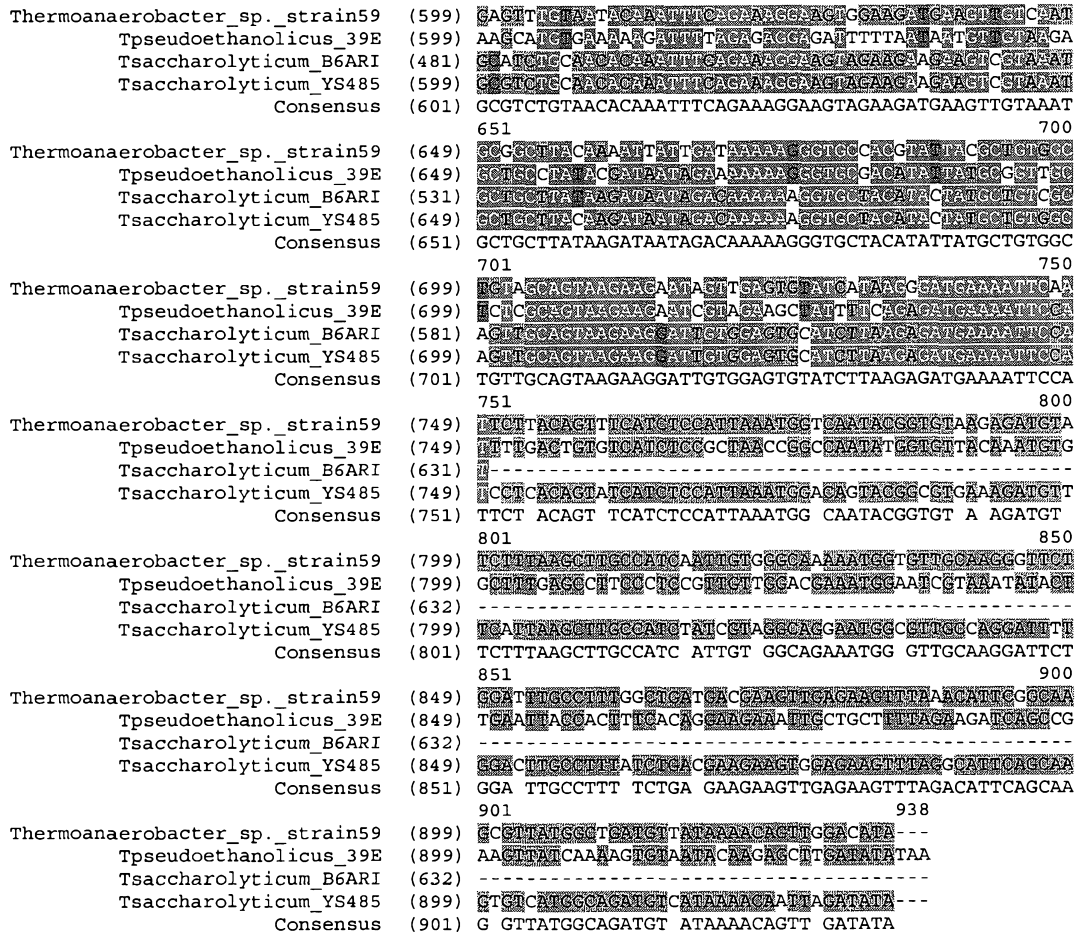

Thermoanaerobacter_sp._strain59 (599) GAGTTTCTAATACAAATTTCAGAAAGGAAGTGGAAGATGAACTTCTCAAT
Tpseudoethanolicus_39E (599) AAGCATCTGTAAAAGATTTTACAGAGGAGATTTTTAATAATTCTTGTAAGA
Tsaccharolyticum_B6ARI (481) GGATCTTCCAACACAATTTCAGAAAGGAAGTAGAAGAAGAACTCCTAAAT
Tsaccharolyticum_YS485 (599) CGGTCTTCCAACACAATTTCAGAAAGGAAGTAGAAGAAGAACTCCTAAAT
Consensus (601) GCGTCTGTAACACAAATTTCAGAAAGGAAGTAGAAGATGAAGTTGTAAAT
651 700
Thermoanaerobacter_sp._strain59 (649) GCGGCTTACTAAATTATTGTTAAAAAGCGTGGCAGGTATTACGCTGTGGC
Tpseudoethanolicus_39E (649) GCTGCCTATTACGATATAGAAAAAAGCGTGGACATATTATGCGGTTGC
Tsaccharolyticum_B6ARI (531) GCTGCTTATTAAGATAATAGACAAAAAAGGTGCTACATATTATGCTGTCGC
Tsaccharolyticum_YS485 (649) GCTGCTTATCTAGATAATAGACAAAAAAGGTGCTACATATTATGCTGTGGC
Consensus (651) GCTGCTTATAAGATAATAGACAAAAAGGGTGCTACATATTATGCTGTGGC
701 750
Thermoanaerobacter_sp._strain59 (699) TGTAGCAGTAAGAAGGATTGTGGAGTGTATCTTAAGAGATGAAAATTCCAA
Tpseudoethanolicus_39E (699) TGTTGCAGTAAGAAGGATTGTGGAGTGTATCTTAAGAGATGAAAATTCCAA
Tsaccharolyticum_B6ARI (581) AGTTGCAGTAAGAAGGATTGTGGAGTGCATCTTAAGAGATGAAAATTCCA
Tsaccharolyticum_YS485 (699) AGTTGCAGTAAGAAGGATTGTGGAGTGCATCTTAAGAGATGAAAATTCCA
Consensus (701) TGTTGCAGTAAGAAGGATTGTGGAGTGTATCTTAAGAGATGAAAATTCCA
751 800
Thermoanaerobacter_sp._strain59 (749) TTCTTACAGTTTCATCTCCATTAAATGGTCAATACGGTGTAAGAGATGTA
Tpseudoethanolicus_39E (749) TTTTGACTGTGTCATCTCCGCTAACCGGCCAATATGGTGTTACAAATGTG
Tsaccharolyticum_B6ARI (631) T----------
Tsaccharolyticum_YS485 (749) TCCTCACAGTATCATCTCCATTAAATGGACAGTACGGCGTCGAAGATGTT
Consensus (751) TTCT ACAGT TCATCTCCATTAAATGG CAATACGGTGT A AGATGT
801 850
Thermoanaerobacter_sp._strain59 (799) TCTTTAAGCTTGCCATCAATTGTGGCAAAAATGGTGTTGCAAGGGTTCT
Tpseudoethanolicus_39E (799) GTTTTGAGCCTTCCCTGCGTTGTTGGACGAAATGGAATCGTAAATATACT
Tsaccharolyticum_B6ARI (632) ----------
Tsaccharolyticum_YS485 (799) TCATTAAGCTTGCCATGTATCCTAGGCAGGAATGGCGTTGCCAGGATTTT
Consensus (801) TCTTTAAGCTTGCCATC ATTGT GGCAGAAATGG GTTGCAAGGATTCT
851 900
Thermoanaerobacter_sp._strain59 (849) GGATTTGCCTTTGGCTCATGACCAAGTTCAGAAGCTTTAAACATTGGGCAA
Tpseudoethanolicus_39E (849) TCAATTACGACTTTCCACAGCAAGAAATTGCTGCTTTTAGAAGATCAGCCG
Tsaccharolyticum_B6ARI (632) ----------
Tsaccharolyticum_YS485 (849) GGACTTGCCTTTATCTGACGAAGAAGTGGAGAAGTTTAGGCATTCAGCAA
Consensus (851) GGA TTGCCTTT TCTGA GAAGAAGTTGAGAAGTTTAGACATTCAGCAA
901 938
Thermoanaerobacter_sp._strain59 (899) GCGTTATGGCTGATGTTATAAAACAGTTGGACATA---
Tpseudoethanolicus_39E (899) AACTTATCAAAAGTGTAATACAAGAGCTTGATATATAA
Tsaccharolyticum_B6ARI (632) ----------
Tsaccharolyticum_YS485 (899) GTGTCATGCGACATGTCATAAAACAATTACATATA---
Consensus (901) G GTTATGGCAGATGT ATAAAACAGTT GATATA

Figure 34

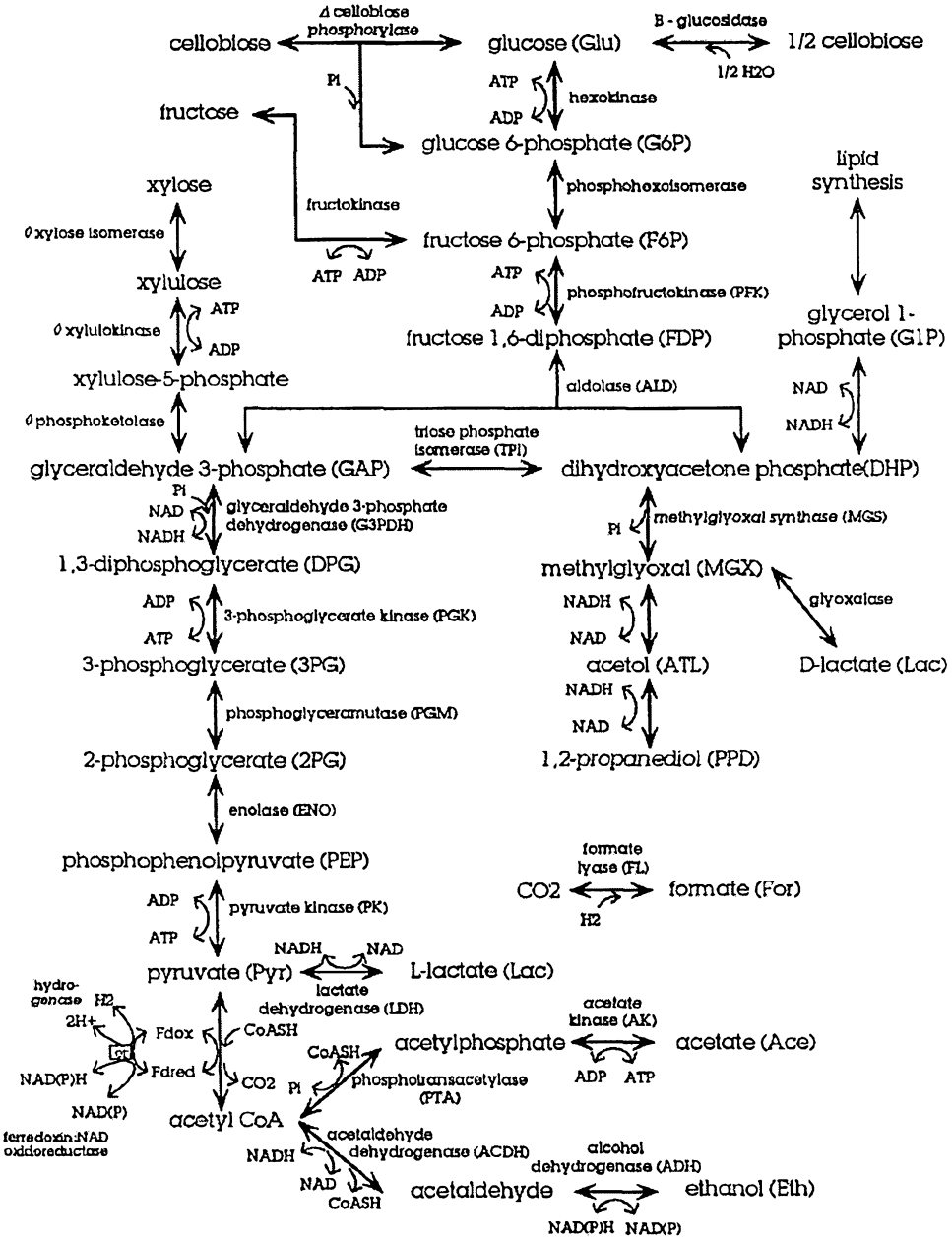

Δ cellobiose phosphorylase
cellobiose
glucose (Glu)
B - glucosidase
1/2 cellobiose
1/2 H2O
Pi
ATP
ADP
hexokinase
fructose
glucose 6-phosphate (G6P)
lipid synthesis
xylose
phosphohexoisomerase
fructokinase
ATP ADP
fructose 6-phosphate (F6P)
Ø xylose isomerase
xylulose
ATP
ADP
phosphofructokinase (PFK)
glycerol 1-phosphate (G1P)
Ø xylulokinase
fructose 1,6-diphosphate (FDP)
xylulose-5-phosphate
aldolase (ALD)
NAD
NADH
Ø phosphoketolase
triose phosphate isomerase (TPI)
glyceraldehyde 3-phosphate (GAP)
dihydroxyacetone phosphate(DHP)
Pi
NAD
NADH
glyceraldehyde 3-phosphate dehydrogenase (G3PDH)
methylglyoxal synthase (MGS)
1,3-diphosphoglycerate (DPG)
methylglyoxal (MGX)
glyoxalase
ADP
ATP
3-phosphoglycerate kinase (PGK)
NADH
NAD
3-phosphoglycerate (3PG)
acetol (ATL)
D-lactate (Lac)
phosphoglyceramutase (PGM)
2-phosphoglycerate (2PG)
1,2-propanediol (PPD)
enolase (ENO)
phosphophenolpyruvate (PEP)
formate lyase (FL)
CO2
formate (For)
H2
ADP
ATP
pyruvate kinase (PK)
NADH
NAD
pyruvate (Pyr)
L-lactate (Lac)
lactate dehydrogenase (LDH)
hydro-genase
H2
2H+
Fdox
CoASH
acetate kinase (AK)
acetylphosphate
acetate (Ace)
CoASH
NAD(P)H
Fdred
CO2
Pi
ADP
ATP
phosphotransacetylase (PTA)
NAD(P)
acetyl CoA
ferredoxin:NAD oxidoreductase
acetaldehyde dehydrogenase (ACDH)
alcohol dehydrogenase (ADH)
NADH
NAD
CoASH
acetaldehyde
ethanol (Eth)
NAD(P)H
NAD(P)

Figure 35

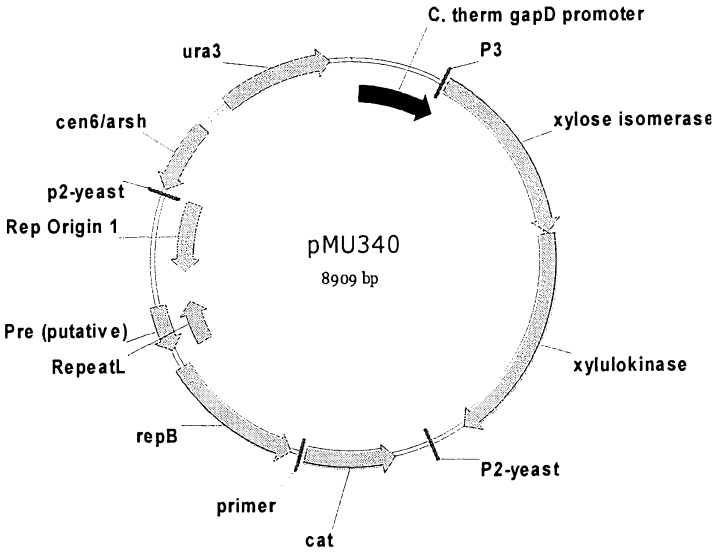

C. therm gapD promoter
P3
ura3
xylose isomerase
cen6/arsh
p2-yeast
Rep Origin 1
pMU340
8909 bp
Pre (putative)
RepeatL
xylulokinase
repB
P2-yeast
primer
cat

Figure 36

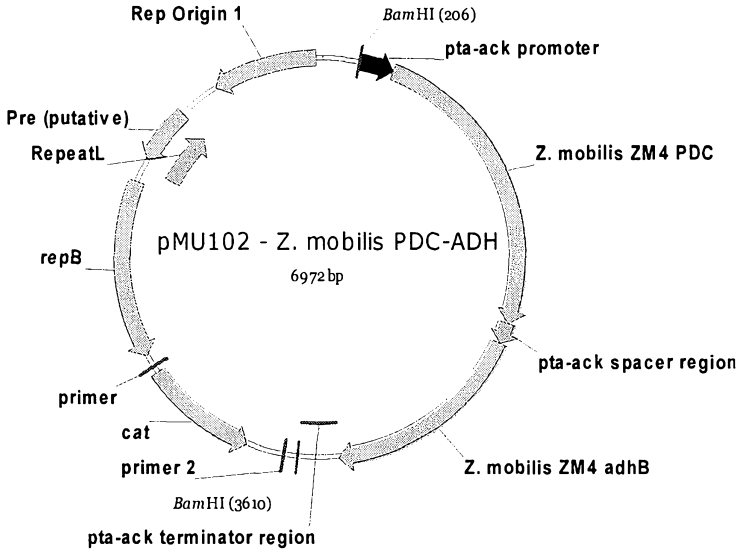

Rep Origin 1
BamHI (206)
pta-ack promoter
Pre (putative)
RepeatL
Z. mobilis ZM4 PDC
pMU102 - Z. mobilis PDC-ADH
6972 bp
repB
pta-ack spacer region
primer
cat
primer 2
Z. mobilis ZM4 adhB
BamHI (3610)
pta-ack terminator region

Figure 37

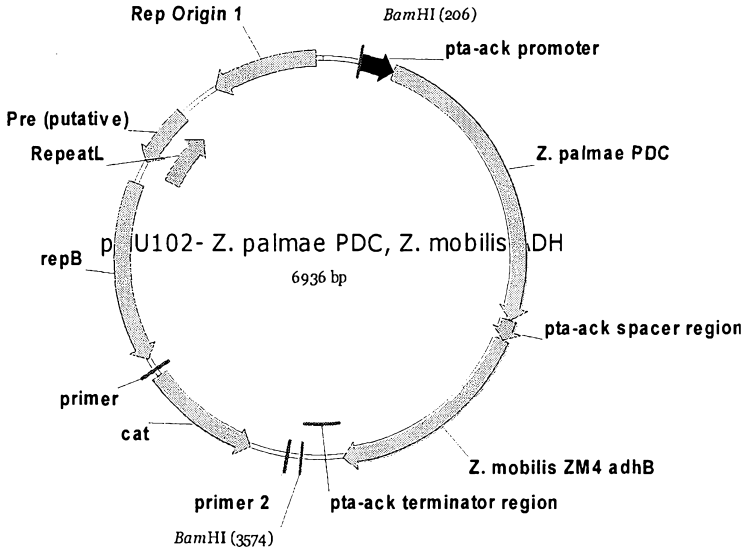

Rep Origin 1
BamHI (206)
pta-ack promoter
Pre (putative)
RepeatL
Z. palmae PDC
repB
pUU102- Z. palmae PDC, Z. mobilis ADH
6936 bp
pta-ack spacer region
primer
cat
Z. mobilis ZM4 adhB
primer 2
pta-ack terminator region
BamHI (3574)

Figure 38

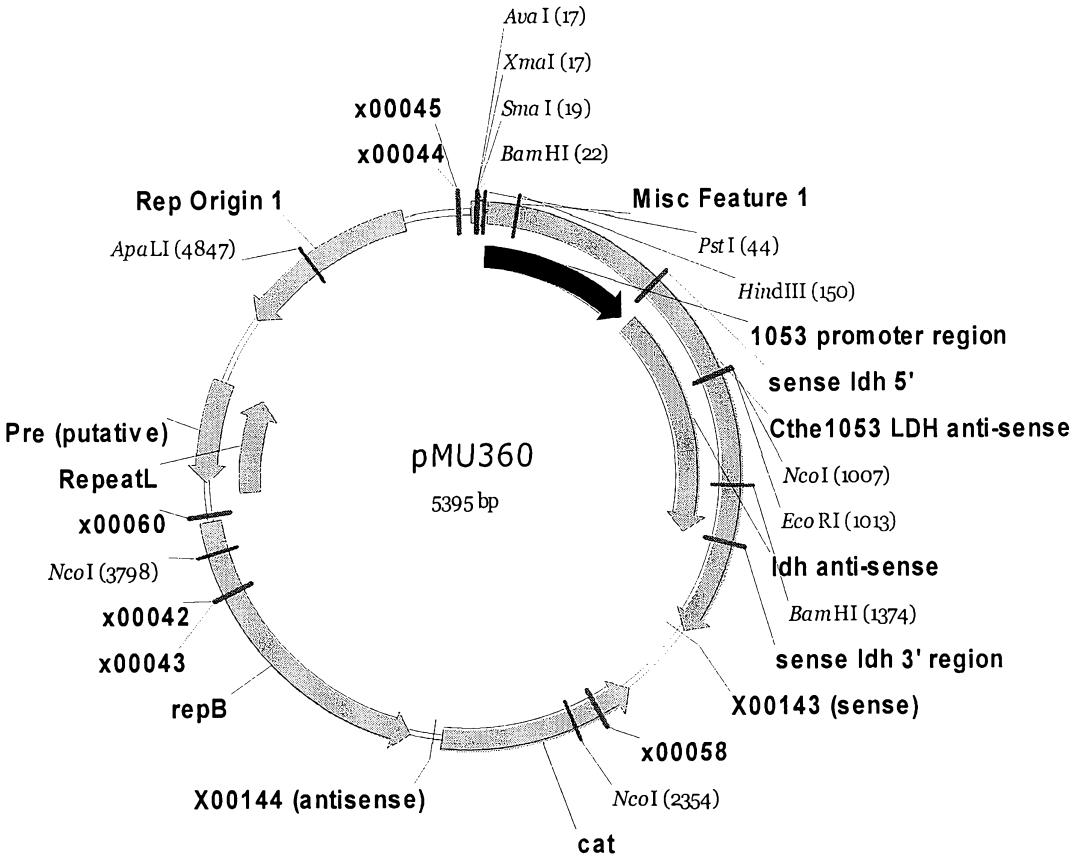

pMU360
5395 bp
AvaI (17)
XmaI (17)
SmaI (19)
BamHI (22)
x00045
x00044
Rep Origin 1
ApaLI (4847)
Misc Feature 1
PstI (44)
HindIII (150)
1053 promoter region
sense ldh 5'
Cthe1053 LDH anti-sense
NcoI (1007)
EcoRI (1013)
ldh anti-sense
BamHI (1374)
sense ldh 3' region
X00143 (sense)
x00058
NcoI (2354)
cat
X00144 (antisense)
repB
x00043
x00042
NcoI (3798)
x00060
RepeatL
Pre (putative)

Figure 39

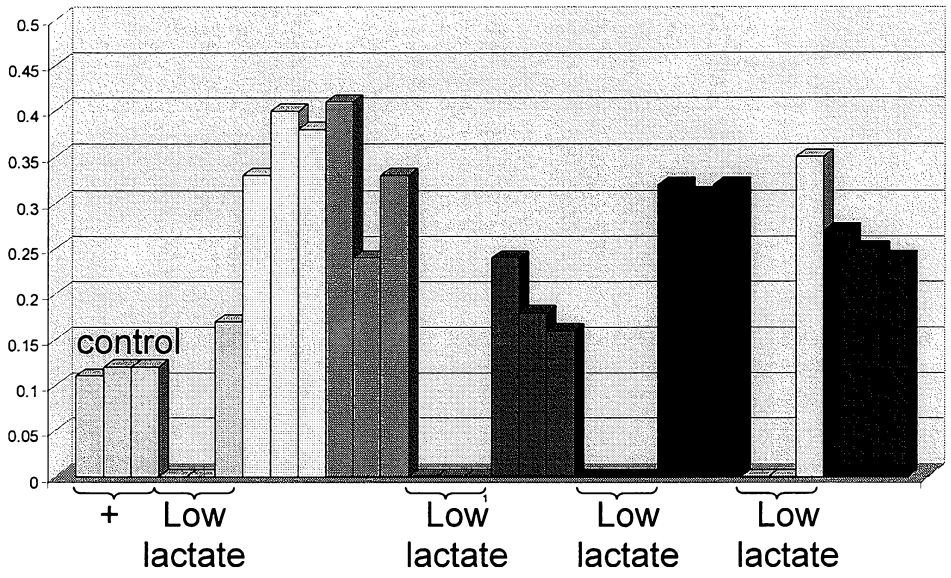

0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
control
+
Low lactate
Low lactate
Low lactate
Low lactate

Figure 40

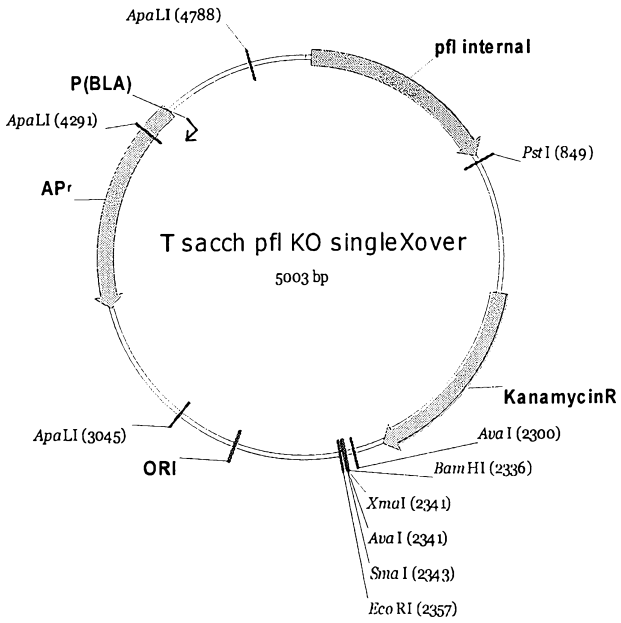

ApaLI (4788)
pfl internal
P(BLA)
ApaLI (4291)
PstI (849)
APr
T sacch pfl KO singleXover
5003 bp
KanamycinR
ApaLI (3045)
AvaI (2300)
ORI
BamHI (2336)
XmaI (2341)
AvaI (2341)
SmaI (2343)
EcoRI (2357)

Figure 41

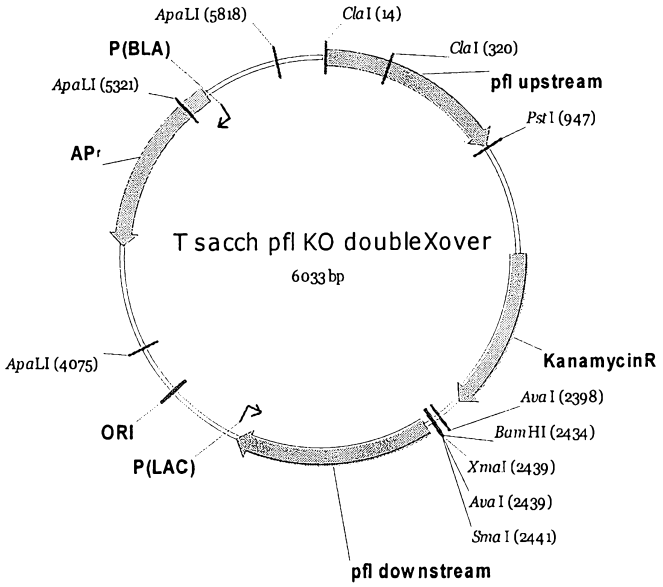

ApaLI (5818)
ClaI (14)
P(BLA)
ClaI (320)
pfl upstream
ApaLI (5321)
PstI (947)
APr
T sacch pfl KO doubleXover
6033 bp
KanamycinR
ApaLI (4075)
AvaI (2398)
BamHI (2434)
ORI
P(LAC)
XmaI (2439)
AvaI (2439)
SmaI (2441)
pfl downstream

Figure 42

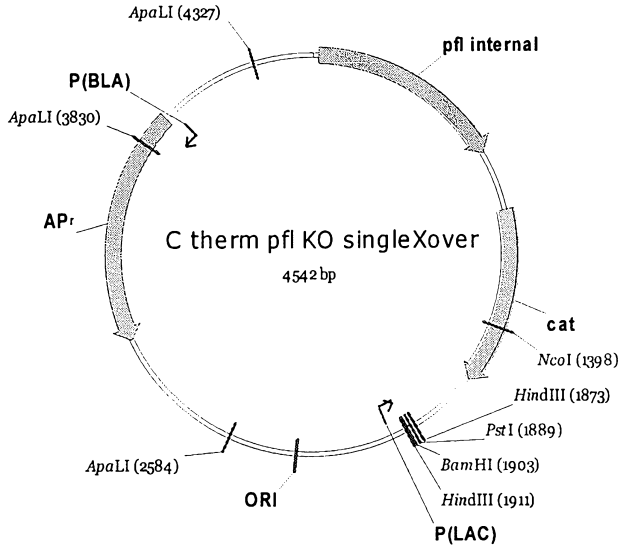

ApaLI (4327)
pfl internal
P(BLA)
ApaLI (3830)
APr
C therm pfl KO singleXover
4542 bp
cat
NcoI (1398)
HindIII (1873)
PstI (1889)
BamHI (1903)
HindIII (1911)
P(LAC)
ApaLI (2584)
ORI

Figure 43

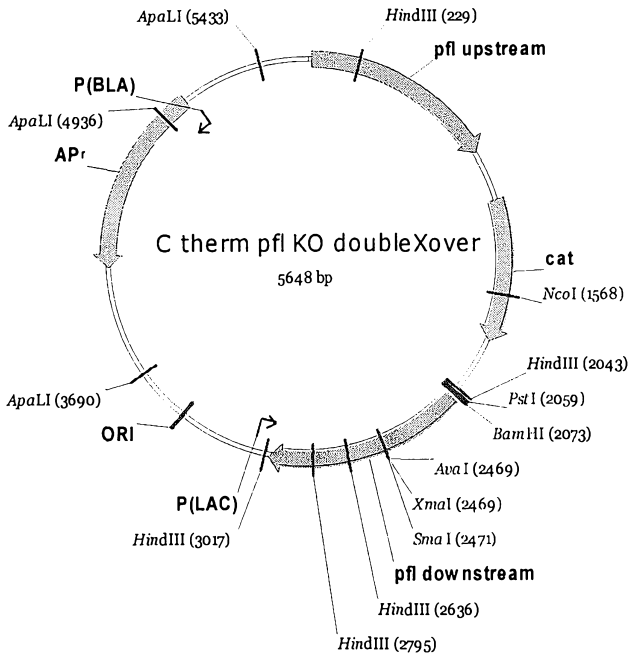

ApaLI (5433)
HindIII (229)
pfl upstream
P(BLA)
ApaLI (4936)
APr
C therm pfl KO doubleXover
5648 bp
cat
NcoI (1568)
HindIII (2043)
PstI (2059)
BamHI (2073)
ApaLI (3690)
ORI
AvaI (2469)
XmaI (2469)
SmaI (2471)
P(LAC)
HindIII (3017)
pfl downstream
HindIII (2636)
HindIII (2795)

Figure 44

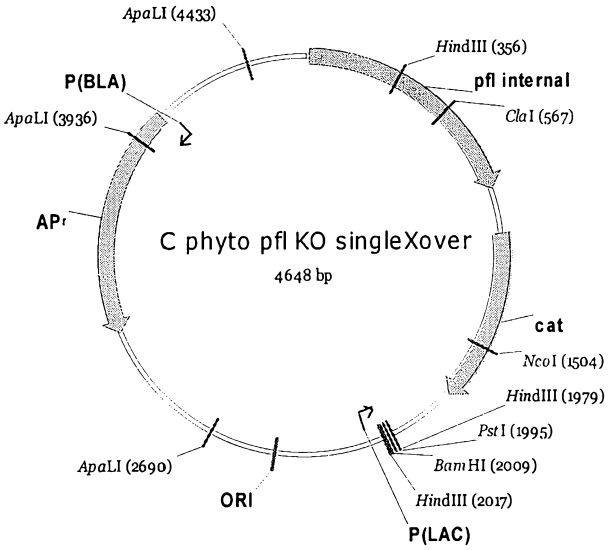

ApaLI (4433)
HindIII (356)
P(BLA)
pfl internal
ApaLI (3936)
ClaI (567)
APr
C phyto pfl KO singleXover
4648 bp
cat
NcoI (1504)
HindIII (1979)
PstI (1995)
ApaLI (2690)
BamHI (2009)
HindIII (2017)
ORI
P(LAC)

Figure 45

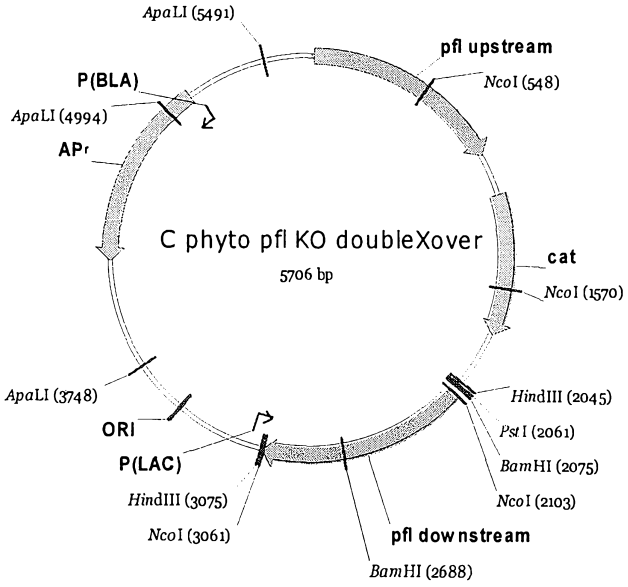

ApaLI (5491)
pfl upstream
NcoI (548)
P(BLA)
ApaLI (4994)
APr
C phyto pfl KO doubleXover
5706 bp
cat
NcoI (1570)
HindIII (2045)
PstI (2061)
BamHI (2075)
NcoI (2103)
ApaLI (3748)
ORI
P(LAC)
HindIII (3075)
NcoI (3061)
pfl downstream
BamHI (2688)

Figure 46

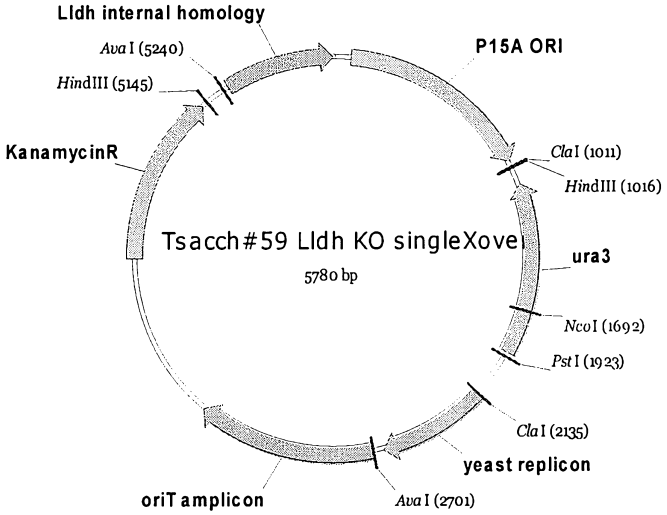

Ldh internal homology
AvaI (5240)
HindIII (5145)
P15A ORI
KanamycinR
ClaI (1011)
HindIII (1016)
Tsacch#59 Ldh KO singleXover
5780 bp
ura3
NcoI (1692)
PstI (1923)
ClaI (2135)
yeast replicon
oriT amplicon
AvaI (2701)

Figure 47

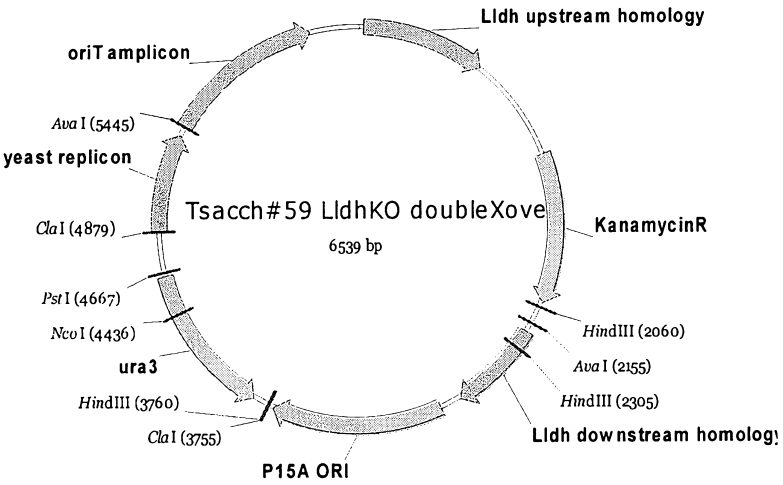

Tsacch#59 LldhKO doubleXove
6539 bp
Ldh upstream homology
oriT amplicon
Ava I (5445)
yeast replicon
Cla I (4879)
Pst I (4667)
Nco I (4436)
ura3
Hind III (3760)
Cla I (3755)
P15A ORI
KanamycinR
Hind III (2060)
Ava I (2155)
Hind III (2305)
Ldh downstream homolog

Figure 48

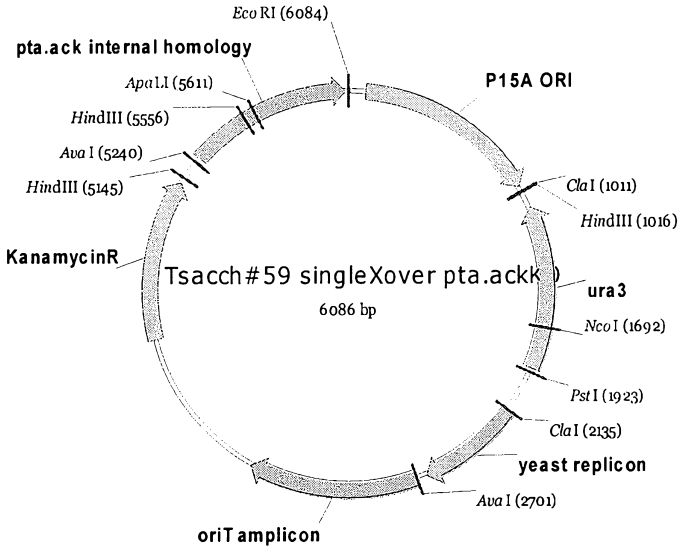

EcoRI (6084)
pta.ack internal homology
ApaLI (5611)
HindIII (5556)
AvaI (5240)
HindIII (5145)
P15A ORI
ClaI (1011)
HindIII (1016)
KanamycinR
Tsacch#59 singleXover pta.ackk
6086 bp
ura3
NcoI (1692)
PstI (1923)
ClaI (2135)
yeast replicon
AvaI (2701)
oriT amplicon

Figure 49

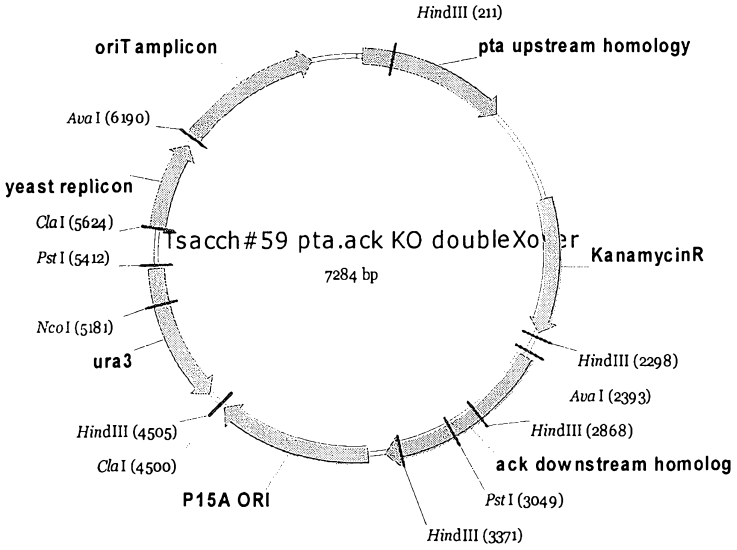

HindIII (211)
pta upstream homology
oriT amplicon
AvaI (6190)
yeast replicon
ClaI (5624)
PstI (5412)
NcoI (5181)
ura3
HindIII (4505)
ClaI (4500)
P15A ORI
HindIII (3371)
PstI (3049)
ack downstream homolog
HindIII (2868)
AvaI (2393)
HindIII (2298)
KanamycinR
Tsacch#59 pta.ack KO doubleXo er
7284 bp

Figure 50

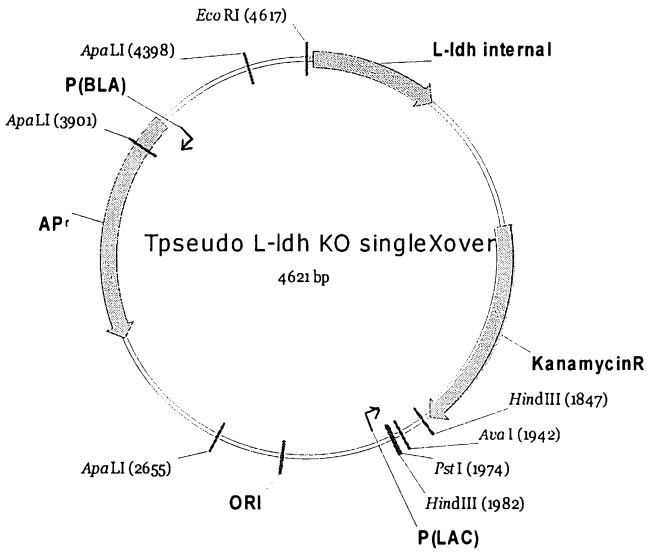

EcoRI (4617)
ApaLI (4398)
L-ldh internal
P(BLA)
ApaLI (3901)
APr
Tpseudo L-ldh KO singleXover
4621 bp
KanamycinR
HindIII (1847)
AvaI (1942)
PstI (1974)
HindIII (1982)
ApaLI (2655)
ORI
P(LAC)

Figure 51

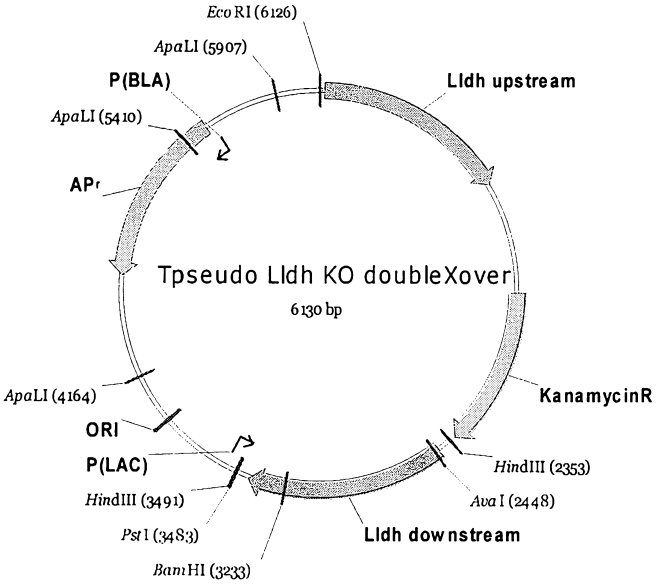

EcoRI (6126)
ApaLI (5907)
P(BLA)
Lldh upstream
ApaLI (5410)
APr
Tpseudo Lldh KO doubleXover
6130 bp
ApaLI (4164)
KanamycinR
ORI
P(LAC)
HindIII (2353)
HindIII (3491)
AvaI (2448)
PstI (3483)
Lldh downstream
BamHI (3233)

Figure 52

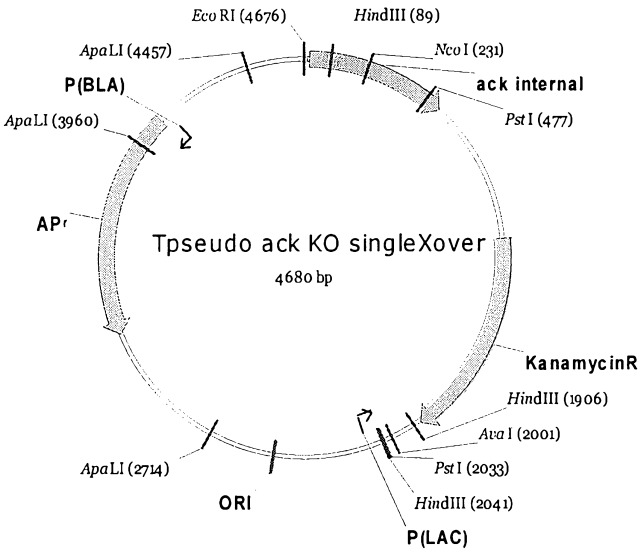
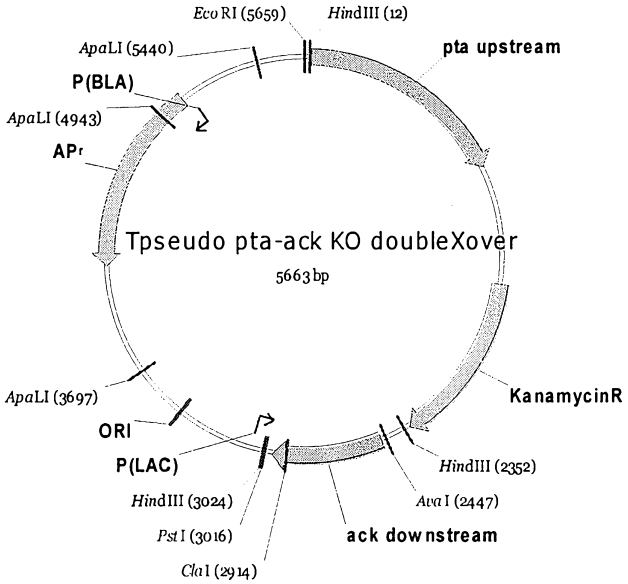

EcoRI (4676)
HindIII (89)
ApaLI (4457)
NcoI (231)
P(BLA)
ack internal
ApaLI (3960)
PstI (477)
APr
Tpseudo ack KO singleXover
4680 bp
KanamycinR
HindIII (1906)
AvaI (2001)
PstI (2033)
ApaLI (2714)
ORI
HindIII (2041)
P(LAC)

Figure 53


EcoRI (5659)
HindIII (12)
ApaLI (5440)
pta upstream
P(BLA)
ApaLI (4943)
APr
Tpseudo pta-ack KO doubleXover
5663 bp
ApaLI (3697)
KanamycinR
ORI
P(LAC)
HindIII (2352)
HindIII (3024)
AvaI (2447)
PstI (3016)
ack downstream
ClaI (2914)