



(86) Date de dépôt PCT/PCT Filing Date: 2004/04/28
(87) Date publication PCT/PCT Publication Date: 2004/11/11
(45) Date de délivrance/Issue Date: 2012/07/03
(85) Entrée phase nationale/National Entry: 2005/10/24
(86) N° demande PCT/PCT Application No.: US 2004/013258
(87) N° publication PCT/PCT Publication No.: 2004/097001
(30) Priorité/Priority: 2003/04/29 (US60/466,831)

(51) Cl.Int./Int.Cl. *C12N 15/56* (2006.01),
A23K 1/00 (2006.01), *C07H 21/00* (2006.01),
C11D 3/30 (2006.01), *C11D 3/386* (2006.01),
C12N 15/31 (2006.01), *C12N 9/42* (2006.01),
C12P 19/14 (2006.01), *C12P 7/08* (2006.01),
C12Q 1/34 (2006.01), *C40B 30/08* (2006.01)

(72) Inventeurs/Inventors:
JONES, BRIAN E., NL;
GRANT, WILLIAM D., GB;
HEAPHY, SHAUN, GB;
GRANT, SUSAN, GB

(73) Propriétaire/Owner:
GENENCOR INTERNATIONAL, INC., US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L., S.R.L.

(54) Titre : CELLULASE 029CEL DE BACILLE ISOLEE
(54) Title: ISOLATED BACILLUS 029 CEL CELLULASE

The translated 029cel protein composed of 581 amino acids

MNFSLPPIRFF	GIGPTAGFFL	LLMGSSALVA	STHTDKIRID	QFGYPADATK	50
VAVIADPQIG	WNSAESYSPG	ATLEVRRVND	GVVVFSGSPV	PWNGGAIHIQ	100
SGDRVWWFDF	TVVAEPGHYR	IHDPANNTHS	DSFAIGADVY	DVVLREAVRM	150
FFYQSRGFAK	EVPIAHANWA	DAASHPQDVA	SRPIWDMGNA	SLERDLSGGW	200
FDAGDFNKYS	EWTEGRVILEL	LLAYQGRPDV	FTDDFGIPES	GNGVPDLLDE	250
VKWGMDWLLR	MQEPSGAILG	KVSVTGHQSA	SPPSTDTHPR	YYGPVSTEAT	300
AMAAAAFALG	ATVFESVGMS	DYAVTLESAA	IAAWNWTMVH	PHVPFDNTGF	350
ASVSPSRNAH	DTLANRVMAA	AMLFERTGGA	VYRDFEDVRY	LDMEPVQWWY	400
FFPFQGELQK	ALAHYTTLPG	ATPSVSADIR	NRMAASINGG	EFLGAWNNQT	450
DAYRAYLKDQ	DYTWGSNKTK	SQAGFFFEGV	RRLGLNPADA	AAHRDAAMGY	500
LHYLHGVNPM	GMVYLSNMYA	SGADRAANEI	YHHWFRDGRT	GTMPSLHSTV	550
PLLVFFRAGP	NAQIQRKHSG	DPRPTRA KSL	P		581

(57) Abrégé/Abstract:

The present invention provides a novel cellulase nucleic acid sequence, designated 029cel, and the corresponding 029cel amino acid sequence. The invention also provides expression vectors and host cells comprising a nucleic acid sequence encoding 029cel, recombinant 029cel proteins and methods for producing the same.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
11 November 2004 (11.11.2004)

PCT

(10) International Publication Number
WO 2004/097001 A2

- (51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US2004/013258

(22) International Filing Date: 28 April 2004 (28.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/466,831 29 April 2003 (29.04.2003) US

(71) Applicant (for all designated States except US): **GENENCOR INTERNATIONAL, INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JONES, Brian, E.** [GB/NL]; Gravin Juliana van Stolberglaan 24, NL-2263VA Leidschendam (NL). **GRANT, William, D.** [GB/GB]; 15 The Tithings, Kibwoth Beauchamp, Leicestershire LE8 OPU (GB). **HEAPHY, Shaun** [GB/GB]; 426 Uppingham Road, Leicester, Leicestershire LE52DP (GB). **GRANT, Susan** [GB/GB]; 15 The Tithings, Kibworth Beauchamp, Leicestershire LE8 OPU (GB).
- (74) Agent: **BOYD, Victoria, L.**; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, HU, IE, IT, LU, MC, NL, 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:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL BACILLUS 029CEL CELLULASE

The translated 029cel protein composed of 581 amino acids

MNFSLP	IRFF	GIGPTAGFFL	LLMGSSALVA	STHTDKIRID	QFGYPADATK	50
VAVIADPQIG	WNSAESYSPG	ATLEVRRVND	GVVVFSGSPV	PWNGGAIHIQ		100
SGDRVWWFDF	TVVAEPGHYR	IHDPANNTHS	DSFAIGADVY	DVVLREAVRM		150
FFYQRS	GFAK	EVPIAHANWA	DAASHPQDVA	SRPIWDMGNA	SLERDLSSGW	200
FDAGDFNKYS	EW TGRVILEL	LLAYQGRPDV	FTDDEGIPES	GNGVPDLLDE		250
VKWGMDWLLR	MQEP SGAILG	KVSVTGHQSA	SPPSTDTHPR	YYGPFVSTEAT		300
AMAAAAFALG	ATVFESVGMS	DYAVTLESAA	IAAWNWTMVH	PHVPFDNTGF		350
ASVSPSRNAH	DTLANRVMAA	AMLFERTGGA	VYRDFFDVRY	LDMEPVQWWY		400
FFPFQGELQK	ALAHYTTLPG	ATPSVSADIR	NRMAASINGG	EFLGAWNNQT		450
DAYRAYLKDQ	DYTWGSNKTK	SQAGFFFE	GGV	RRLGLNPADA	AAHRDAAMGY	500
LHYLHGVNPM	GMVYLSNMYA	SGADRAANEI	YHHWFRDGRT	GTMPSLHSTV		550
PLLVFFRAGP	NAQIQRKHSG	DPRPTRAKSL	P			581

(57) Abstract: The present invention provides a novel cellulase nucleic acid sequence, designated 029cel, and the corresponding 029cel amino acid sequence. The invention also provides expression vectors and host cells comprising a nucleic acid sequence encoding 029cel, recombinant 029cel proteins and methods for producing the same.

WO 2004/097001 A2

ISOLATED BACILLUS 029 CEL CELLULASE

[01]

5

**STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY
SPONSORED RESEARCH AND DEVELOPMENT**

[02] Not applicable.

10

FIELD OF THE INVENTION

15

[03] This invention relates to a novel cellulase referred to herein as 029cel. Also described are nucleic acids encoding the cellulase, compositions comprising said cellulase, methods of identifying novel cellulases and methods of using said compositions. Preferably the cellulase(s) are isolated from *Bacillus* species, preferably *B. agaradhaerens*. The present invention further relates to the use of the novel cellulase in compositions recognized in the art as advantageously having cellulase added thereto, including, as an additive in a detergent composition, in the treatment of cellulose containing fabrics, in the treatment of pulp and paper and in the treatment of starch for the production of high fructose corn-syrup or ethanol.

20

BACKGROUND OF THE INVENTION

25

[04] Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms, including bacteria, yeast and fungi, that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro *et al.*, 2001). As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna *et al.*, 2001). The effective utilization of cellulose through biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya *et al.*, 1997).

30

[05] Cellulases are enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, celooligosaccharides, and the like. Cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91)

("CBH") and beta-glucosidases ([beta] -D-glucoside glucohydrolase; EC 3.2.1.21) ("BG"). (Knowles *et al.*, 1987; Shulein, 1988). Endoglucanases act mainly on the amorphous parts of the cellulose fibre, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, 1995). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki, *et al.* 2000). Beta-glucosidase acts to liberate D-glucose units from cellobiose, cello-oligosaccharides, and other glucosides (Freer, 1993).

[06] In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho *et al.*, 1996). A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH- type cellulases synergistically interact to more efficiently degrade cellulose. See, e.g., Wood, 1985.

[07] Although cellulase compositions have been previously described, there remains a need for new and improved cellulase compositions for use in household detergents, stonewashing compositions or laundry detergents, etc. Cellulases that exhibit improved performance are of particular interest.

BRIEF SUMMARY OF THE INVENTION

[08] It is an object of the present invention to provide a novel cellulase having beneficial properties for use in detergents, treating textiles, biomass conversion and pulp and paper manufacturing.

[09] It is an object of the present invention to provide polypeptides having cellulolytic activity and polynucleotides encoding the polypeptides. The polypeptides may improve the degradation of cell wall material, e.g., cellulose and/or hemicellulose. The polypeptides may also improve the stability or activity of other enzymes involved in the degradation of plant cell wall material, e.g., biomass.

[10] An object of the present invention is to provide a novel cellulase and derivatives thereof, methods of producing such cellulases, and compositions comprising such novel cellulases. The present invention further relates to the use of the novel cellulase and derivatives thereof in compositions recognized in the art as advantageously having cellulase added thereto, including, as an additive in a detergent composition, in the treatment of textiles such as cellulose-containing fabrics and fibers useful therefor, as an animal feed additive, in biomass conversion, in the treatment of pulp and paper and in the treatment of starch for the production of high fructose corn-syrup or ethanol.

[11] It is a further object of the present invention to provide for a method of producing a novel cellulase via heterologous expression from recombinant host cells.

[12] It is yet a further object of the present invention to provide a nucleic acid sequence encoding the inventive cellulase. In one aspect, the nucleic acid and amino acid sequence facilitate commercial production of the novel cellulase and cellulase compositions of the invention.

5 [13] It is still a further object of the present invention to provide a novel cellulase having excellent properties for use in detergents, treating textiles, as a feed supplement and in pulp and paper manufacturing. In a further aspect, the cellulase finds use in biomass conversion.

[14] In a first aspect, the invention includes an isolated polynucleotide having a sequence which encodes 029cel, a sequence complementary to the *029cel* gene coding sequence,
10 and a composition comprising the polynucleotide. The polynucleotide may be mRNA, DNA, cDNA, genomic DNA, or an antisense analog thereof.

[15] In one embodiment, a *029cel* polynucleotide may comprise an isolated nucleic acid molecule which hybridizes to the complement of the nucleic acid presented as SEQ ID NO:2 under moderate to high stringency conditions, where the nucleic acid molecule encodes a
15 *029cel* polypeptide that exhibits cellulose binding activity.

[16] The polynucleotide having at least 80%, 85%, 90%, 95%, 98% or more sequence identity to the sequence presented as SEQ ID NO:2 may encode a *029cel* protein. In a specific embodiment, the polynucleotide comprises a sequence substantially identical to SEQ ID NO:2. The invention also contemplates fragments of the polynucleotide, preferably
20 at least about 15-30 nucleotides in length.

[17] In a second aspect, a novel cellulase or a derivative is provided which is obtainable from a *Bacillus*. Preferably, the cellulase of the invention comprises an amino acid sequence according to Figure 3 (SEQ ID NO:3), a fragment, or a derivative thereof, having greater than 90% sequence identity, preferably greater than 95% sequence identity and
25 more preferably greater than 97% sequence identity to an active portion thereto.

[18] In a third aspect the present invention relates to a nucleic acid construct comprising the nucleotide sequence, which encodes for the polypeptide of the invention, operably linked to one or more control sequences that direct the production of the polypeptide in a suitable host.

30 [19] The invention further provides recombinant expression vectors containing a nucleic acid sequence encoding *029cel* or a fragment or splice variant thereof, operably linked to regulatory elements effective for expression of the protein in a selected host. In a related aspect, the invention includes a host cell containing the vector.

[20] In a fourth aspect the present invention relates to a recombinant expression vector
35 comprising the nucleic acid construct of the invention.

[21] In a fifth aspect the present invention relates to a recombinant host cell comprising the nucleic acid construct of the invention.

[22] The invention further includes a method for producing 029cel by recombinant techniques, by culturing recombinant prokaryotic or eukaryotic host cells comprising nucleic acid sequence encoding 029cel under conditions effective to promote expression of the protein, and subsequent recovery of the protein from the host cell or the cell culture medium.

5 [23] In a sixth aspect the present invention relates to a method for producing a polypeptide of the invention, the method comprising: (a) cultivating a microorganism, which in its wild-type form is capable of producing the polypeptide, to produce the polypeptide; and (b) recovering the polypeptide.

10 [24] In a seventh aspect the invention provides for an enzymatic composition useful in the conversion of cellulose to ethanol. In a preferred embodiment the enzymatic composition comprises 029cel. The composition may further comprise additional cellulase enzymes such as endoglucanases and/or cellbiohydrolases. The composition may be enriched in 029cel.

15 [25] In one embodiment the invention provides a method of identifying novel enzymes by isolating total microbial community DNA from an environment, constructing a genomic DNA library in *E.coli*, screening the library for expression of cellulase activity, identifying the cellulase gene in the cellulase-positive clone and characterising the novel cellulase enzyme.

[26] Further provided herein are analytical methods for detecting 029cel/ nucleic acids and 029cel proteins also form part of the invention.

20 [27] According to yet another embodiment of the invention, a method of transforming a suitable microorganism with nucleic acid sequence encoding a cellulase according to the invention is provided. A method of producing the cellulase according to the invention from said transformed microorganism is provided.

25 [28] A further object of the invention is to provide an expression vector particularly effective in *Streptomyces*. *Streptomyces* serve as alternate host cells for the production of various proteins and with respect to the expression and production of cellulases may offer a number of advantages over *Bacillus* host cells particularly when cells are grown at a high cell density. A preferred expression vector comprises a regulatory polynucleotide sequence including a promoter sequence derived from a glucose isomerase gene of *Actinoplanes*, a signal sequence derived from a *Streptomyces* cellulase gene, and a DNA sequence
30 encoding a cellulase, particularly a cellulase according to the invention.

[29] In a preferred embodiment of the present invention, a full-length cellulase is obtainable from *Bacillus*.

35 [30] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications

within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

- 5 [31] Figure 1 illustrates the environmental nucleotide sequence (SEQ ID NO:1).
[32] Figure 2 illustrates a nucleic acid sequence encoding the novel cellulase (SEQ ID NO:2).
[33] Figure 3 illustrates the deduced amino acid sequence of the inventive cellulase (SEQ ID NO:3).

10 DETAILED DESCRIPTION

[34] The invention will now be described in detail by way of reference only using the following definitions and examples.

15 [35] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a
20 general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences
25 are written left to right in amino to carboxy orientation, respectively. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

[36] The headings provided herein are not limitations of the various aspects or
30 embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[37]

I. DEFINITIONS

[38] "Cellulase," "cellulolytic enzymes" or "cellulase enzymes" means the inventive bacterial endoglucanase described herein. Three different types of cellulase enzymes act synergistically to convert cellulose and its derivatives to glucose.

5 [39] The term "cellulase" refers to a category of enzymes capable of hydrolyzing cellulose polymers to shorter cello-oligosaccharide oligomers, cellobiose and/or glucose. Numerous examples of cellulases, such as exoglucanases, exocellobiohydrolases, endoglucanases, and glucosidases have been obtained from cellulolytic organisms, particularly including fungi, and bacteria. The enzymes made by these microbes are mixtures of proteins with
10 three types of actions useful in the conversion of cellulose to glucose: endoglucanases (EG), cellobiohydrolases (CBH), and beta-glucosidase. These three different types of cellulase enzymes act synergistically to convert cellulose and its derivatives to glucose.

[40] Many microbes make enzymes that hydrolyze cellulose, including the wood rotting fungus *Trichoderma*, the compost bacteria *Thermomonospora*, *Bacillus*, and *Cellulomonas*;
15 *Streptomyces*; and the fungi *Humicola*, *Aspergillus* and *Fusarium*.

[41] By the term "host cell" is meant a cell that contains a vector and supports the replication, and/or transcription or transcription and translation (expression) of the expression construct. Host cells for use in the present invention can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast, plant, insect, amphibian, or mammalian
20 cells. In a one embodiment according to the present invention, "host cell" means the cells of the genus *Bacillus*. In another preferred embodiment according to the invention, "host cell" means the cells of *Streptomyces*. A *Streptomyces* means any bacterial strain that is a member of the genus *Streptomyces* as classified in Buchanan et al., *The Shorter Bergey's Manual For Determinative Bacteriology* (Williams & Wilkens 1982). Particularly preferred
25 strains of *Streptomyces* include *S. lividens*, *S. rubiginosus*, and *S. coelicolor*. *S. lividens* is described in Lomovskaya et al., *J. Virology* 9:258 (1972). However, one of skill will realize that any appropriate host cell, e.g., bacterial, fungal, eukaryotic and plant cell may be used.

[42] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by
30 the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

35 [43] The term "secretory signal sequence" denotes a DNA sequence that encodes a polypeptide (a "secretory peptide" or "secretory signal peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in

which it is synthesized. The larger peptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway to yield the secretory signal peptide and a smaller peptide commonly referred to as the mature polypeptide.

[44] As used herein, the phrases "whole cellulase preparation" and "whole cellulase composition" are used interchangeably and refer to both naturally occurring and non-naturally occurring compositions. A "naturally occurring" composition is one produced by a naturally occurring source and which comprises, for example, one or more cellobiohydrolase-type, one or more endoglucanase-type, and one or more β -glucosidase components wherein each of these components is found at the ratio produced by the source. Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and beta-glucosidases or BG-type cellulases (Schulein, 1988). However, sometimes these systems lack CBH-type cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. A naturally occurring composition is one that is produced by an organism unmodified with respect to the cellulolytic enzymes such that the ratio of the component enzymes is unaltered from that produced by the native organism.

[45] A "non-naturally occurring" composition encompasses those compositions produced by: (1) combining component cellulolytic enzymes either in a naturally occurring ratio or non-naturally occurring, i.e., altered, ratio; or (2) modifying an organism to overexpress or underexpress one or more cellulolytic enzyme; or (3) modifying an organism such that at least one cellulolytic enzyme is deleted or (4) modifying an organism to express a heterologous component cellulolytic enzyme.

[46] As used herein, the term "promoter" refers to a nucleic acid sequence that functions to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences") are necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. The promoter may be the promoter normally associated with the downstream gene or it may be heterologous, i.e., from another gene or another microorganism as long as it function to direct the gene. A preferred promoter when the transformation host cell is *Bacillus* is the *aprE* promoter. In one aspect the promoter is an inducible promoter. In one aspect, when the host cell is a filamentous fungus, the promoter is the *T. reesei* *cbh1* promoter which is deposited in GenBank under Accession Number D86235. In another aspect the promoter is a *cbh II* or xylanase promoter from *T. reesei*.

[47] A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA encoding a secretory leader, i.e., a signal peptide, is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably
5 linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at
10 convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

[48] "DNA construct" or "DNA vector" means a nucleotide sequence which comprises one or more DNA fragments encoding the novel cellulase. Included in "DNA vectors" are "expression vectors." Typical expression vectors contain regulatory sequences such as,
15 transcription and translation terminators, transcription and translation initiation sequences, signal sequences, and promoters useful for regulation of the expression of the particular nucleic acid. The term "promoter" is used in its ordinary sense to refer to a polynucleotide sequence involved in the control of the initiation of transcription of a polynucleotide sequence encoding a protein. A "signal sequence" refers to a signal peptide or a portion of a
20 protein that is capable of directing the transport of a desired protein in bioactive form from a host. The mature form of an extracellular protein lacks the signal sequence which is cleaved off during the secretion process. While not meant to limit the invention, the number of amino acid residues in a signal peptide may be between about 5 and about 100 amino acid residues. Signal sequence may be modified to provide for cloning sites that allow for the
25 ligation of DNA or insertion of DNA encoding a cellulase. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in prokaryotes, eukaryotes, or both, (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or both. See, Gilman and Smith, *Gene* 8:81-97 (1979); Roberts *et al.*, *Nature* 328:731-734 (1987); Berger and
30 Kimmel, *GUIDE TO MOLECULAR CLONING TECHNIQUES, METHODS IN ENZYMOLOGY, VOL 152*, Academic Press, Inc., San Diego, CA ("Berger"); Scheider, B., *et al.*, *Protein Expr. Purif* 6435:10 (1995); Sambrook *et al.* *MOLECULAR CLONING - A LABORATORY MANUAL (2ND ED.) VOL. 1-3*, Cold Springs Harbor Publishing (1989) ("Sambrook"); and *CURRENT PROTOCOLS IN*
35 *MOLECULAR BIOLOGY*, Ausubel *et al.* (eds.), Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1997 Supplement) ("Ausubel"). Cloning vectors useful in *Streptomyces* are known and reference is made to

U.S. Patent Nos. 4,338,397; 4,411,994; 4,513,085; 4,513,086; 4,745,056; 5,514,590; and 5,622,866 and WO88/07079.

[49] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, that may or may not include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[50] The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not normally found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences, e.g., from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein will often refer to two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

[51] The term "% homology" is used interchangeably herein with the term "% identity" herein and refers to the level of nucleic acid or amino acid sequence identity between the nucleic acid sequence that encodes 029cel or the 029cel amino acid sequence, when aligned using a sequence alignment program.

[52] For example, as used herein, 80% homology means the same thing as 80% sequence identity determined by a defined algorithm, and accordingly a homologue of a given sequence has greater than 80% sequence identity over a length of the given sequence. Exemplary levels of sequence identity include, but are not limited to, 80, 85, 90, 95, 98% or more sequence identity to a given sequence, e.g., the coding sequence for 029cel, as described herein.

[53] Exemplary computer programs which can be used to determine identity between two sequences include, but are not limited to, the suite of BLAST programs, e.g., BLASTN, BLASTX, and TBLASTX, BLASTP and TBLASTN.

See also, Altschul, *et al.*, 1990 and Altschul, *et al.*, 1997.

[54] Sequence searches are typically carried out using the BLASTN program when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences that have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other public databases. Both BLASTN and BLASTX are run using default parameters of an open gap penalty of 11.0, and an extended gap penalty of 1.0, and utilize the BLOSUM-62 matrix. (See, e.g., Altschul, *et al.*, 1997.)

[55] A preferred alignment of selected sequences in order to determine "% identity" between two or more sequences, is performed using for example, the CLUSTAL-W program in MacVector version 6.5, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM 30 similarity matrix.

5 [56] The terms "isolated" or "purified" as used herein refer to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

[57] In the present context, the term "substantially pure polypeptide" means a polypeptide preparation which contains at the most 10% by weight of other polypeptide material with which it is natively associated (lower percentages of other polypeptide material are preferred, e.g. at the most 8% by weight, at the most 6% by weight, at the most 5% by weight, at the most 4% at the most 3% by weight, at the most 2% by weight, at the most 1% by weight, and at the most 1/2% by weight). Thus, it is preferred that the substantially pure polypeptide is at least 92% pure, i.e. that the polypeptide constitutes at least 92% by weight of the total polypeptide material present in the preparation, and higher percentages are preferred such as at least 94% pure, at least 95% pure, at least 96% pure, at least 96% pure, at least 97% pure, at least 98% pure, at least 99%, and at the most 99.5% pure. The polypeptides disclosed herein are preferably in a substantially pure form. In particular, it is preferred that the polypeptides disclosed herein are in "essentially pure form", i.e. that the polypeptide preparation is essentially free of other polypeptide material with which it is natively associated. This can be accomplished, for example, by preparing the polypeptide by means of well-known recombinant methods. Herein, the term "substantially pure polypeptide" is synonymous with the terms "isolated polypeptide" and "polypeptide in isolated form".

25 [58] In general, nucleic acid molecules which encode the 029cel will hybridize, under moderate to high stringency conditions to the sequence provided herein as SEQ ID NO:2 (the 029cel). However, in some cases a 029cel-encoding nucleotide sequence is employed that possesses a substantially different codon usage, while the protein encoded by the 029cel-encoding nucleotide sequence has the same or substantially the same amino acid sequence as the native protein. For example, the coding sequence may be modified to facilitate faster expression of 029cel in a particular prokaryotic or eukaryotic expression system, in accordance with the frequency with which a particular codon is utilized by the host. Te'o, *et al.* (2000), for example, describes the optimization of genes for expression in filamentous fungi.

35 [59] A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex or probe. For

example, "maximum stringency" typically occurs at about $T_m - 5^\circ\text{C}$ (5° below the T_m of the probe); "high stringency" at about $5-10^\circ$ below the T_m ; "moderate" or "intermediate stringency" at about $10-20^\circ$ below the T_m of the probe; and "low stringency" at about $20-25^\circ$ below the T_m . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

[60] Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook, *et al*, 1989, Chapters 9 and 11, and in Ausubel, F.M., *et al.*, 1993).

An example of high stringency conditions includes hybridization at about 42°C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 $\mu\text{g/ml}$ denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at 42°C .

[61] As used herein, the terms "transformed", "stably transformed" or "transgenic" with reference to a cell means the cell has a non-native (heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations.

[62] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

[63] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

[64] It follows that the term "029cel expression" refers to transcription and translation of the 029cel cellulase gene, the products of which include precursor RNA, mRNA, polypeptide, post-translationally processed polypeptides. By way of example, assays for 029cel expression include Western blot for 029cel protein, Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for 029cel mRNA, and endoglucanase activity assays as described in Shoemaker S.P. and Brown R.D.Jr. (Biochim. Biophys. Acta, 1978, 523:133-146) and Schulein (1988).

[65] As used herein, the term "surfactant" refers to any compound generally recognized in the art as having surface active qualities. Thus, for example, surfactants comprise anionic, cationic and nonionic surfactants such as those commonly found in detergents. Anionic

surfactants include linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; and alkanesulfonates. Ampholytic surfactants include quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants may comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like.

[66] As used herein, the term "cellulose containing fabric" refers to any sewn or unsewn fabrics, yarns or fibers made of cotton or non-cotton containing cellulose or cotton or non-cotton containing cellulose blends including natural cellulosics and manmade cellulosics (such as jute, flax, ramie, rayon, and lyocell).

[67] As used herein, the term "cotton-containing fabric" refers to sewn or unsewn fabrics, yarns or fibers made of pure cotton or cotton blends including cotton woven fabrics, cotton knits, cotton denims, cotton yarns, raw cotton and the like.

[68] As used herein, the term "stonewashing composition" refers to a formulation for use in stonewashing cellulose containing fabrics. Stonewashing compositions are used to modify cellulose containing fabrics prior to sale, *i.e.*, during the manufacturing process. In contrast, detergent compositions are intended for the cleaning of soiled garments and are not used during the manufacturing process.

[69] As used herein, the term "detergent composition" refers to a mixture which is intended for use in a wash medium for the laundering of soiled cellulose containing fabrics. In the context of the present invention, such compositions may include, in addition to cellulases and surfactants, additional hydrolytic enzymes, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, cellulase activators, antioxidants, and solubilizers.

[70] As used herein, the terms "active" and "biologically active" refer to a biological activity associated with a particular protein and are used interchangeably herein. For example, the enzymatic activity associated with a protease is proteolysis and, thus, an active protease has proteolytic activity. It follows that the biological activity of a given protein refers to any biological activity typically attributed to that protein by those of skill in the art.

[71] When employed in enzymatic solutions, the 029cel component is generally added in an amount sufficient to allow the highest rate of release of soluble sugars from the biomass. The amount of 029cel component added depends upon the type of biomass to be saccharified which can be readily determined by the skilled artisan. However, when employed, the weight percent of the 029cel component relative to any other cellulase type components present in the cellulase composition is from preferably about 1, preferably about

5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 40 weight percent, from about 15 to about 45 weight percent, from about 15 to about 50 weight percent.

II. MOLECULAR BIOLOGY

[72] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Ausubel *et al.*, eds., *Current Protocols in Molecular Biology* (1994)).

[73] To obtain high level expression of a cloned gene, the heterologous gene is preferably positioned about the same distance from the promoter as is in the naturally occurring cellulase gene. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

[74] Those skilled in the art are aware that a natural promoter can be modified by replacement, substitution, addition or elimination of one or more nucleotides without changing its function. The practice of the invention encompasses and is not constrained by such alterations to the promoter.

[75] The expression vector/construct typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the heterologous sequence. A typical expression cassette thus contains a promoter operably linked to the heterologous nucleic acid sequence and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination.

Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

[76] The practice of the invention is not constrained by the choice of promoter in the genetic construct. The only constraint on the choice of promoter is that it is functional in the host cell used. A preferred promoter when the transformation host cell is *Bacillus* is the *aprE* promoter.

[77] In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

[78] The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include bacteriophages λ and M13, as well as plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as MBP, GST, and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, e.g., c-myc.

[79] The elements that are typically included in expression vectors also include a replicon, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of heterologous sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable.

[80] The methods of transformation of the present invention may result in the stable integration of all or part of the transformation vector into the genome of the filamentous fungus. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

[81] The gene encoding the cellulase of the present invention can be cloned using λ -phage (expression) vectors and *E. coli* host cells. (Alternatively PCR cloning using consensus primers designed on conserved domains may be used.) Applicants have discovered that transformation of the gene encoding the cellulase of the present invention and expression in *E. coli* results in an active protein. After a first cloning step in *E. coli*, a cellulase gene according to the present invention can be transferred to a more preferred industrial expression host such as *Bacillus* or *Streptomyces* species, a filamentous fungus such as *Aspergillus* or *Trichoderma*, or a yeast such as *Saccharomyces*. High level expression and secretion obtainable in these host organisms allows accumulation of the cellulase in the fermentation medium from which it can subsequently be recovered.

[82] A preferred general transformation and expression protocol for protease deleted *Bacillus* strains is provided in Ferrari et al., U.S. Patent No. 5,264,366.

Transformation and expression in *Aspergillus* is described in, for example, Berka et al., U.S. Patent No. 5,364,770.

5 [83] Many standard transfection methods can be used to produce *Trichoderma reesei* cell lines that express large quantities of the heterologous protein. Some of the published methods for the introduction of DNA constructs into cellulase-producing strains of *Trichoderma* include Lorito, Hayes, DiPietro and Harman, 1993, Curr. Genet. 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, 1990, Curr. Genet. 17:169-174; Penttila,
10 Nevalainen, Ratto, Salminen and Knowles, 1987, Gene 6: 155-164, for *Aspergillus* Yelton, Hamer and Timberlake, 1984, Proc. Natl. Acad. Sci. USA 81: 1470-1474, for *Fusarium* Bajar, Podila and Kolattukudy, 1991, Proc. Natl. Acad. Sci. USA 88: 8202-8212, for *Streptomyces* Hopwood et al., 1985, The John Innes Foundation, Norwich, UK and for *Bacillus* Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, 1990, FEMS Microbiol. Lett. 55:
15 135-138).

[84] However, any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for
20 introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). Also of use is the *Agrobacterium*-mediated transfection method described in U.S. Patent No. 6,255,115. It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the heterologous gene.

25 [85] After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of genes under control of cellulase gene promoter sequences. Large batches of transformed cells can be cultured as described below. Finally, product is recovered from the culture using standard techniques.

[86] Thus, the invention herein provides for the expression and enhanced secretion of the
30 inventive cellulases whose expression is under control of cellulase gene promoter sequences including naturally occurring cellulase genes, fusion DNA sequences, and various heterologous constructs. The invention also provides processes for expressing and secreting high levels of the inventive cellulases.

35 III. Identification of Nucleic Acids and Encoded Protein Sequences

[87] A genomic library from *Bacillus agaradhaerans* (DSM 8721) was prepared using standard techniques known in the art. This organism produces an alkaline cellulase, (endo-

1,4-beta-glucanase), belonging to cellulase family 5 of glycosyl hydrolases, endoglucanase 5A, EC 3.2.1.4, Swiss-Prot : 085465, entry name GUN5_BACAG. EBI accession number AF067428) the gene for which is 1203bp in length, (Davies et al. 1998). Cellulase positive clones were detected with an incidence of 1/3000 in the plate assay. In the process for isolating a gene according to an aspect of the present invention, degenerate primers based on the coding sequence for this enzyme were used. Unexpectedly, however, no PCR product was obtained using primers known to amplify the known *B. agaradhaerans* cellulase. The complete sequence of the insert coding for the cellulase was therefore determined by primer walking.

10 [88] The process for isolating a gene according to the second aspect of the present invention makes use of its homology to a nucleotide sequence comprising all or part of the nucleotide sequence of SEQ ID No.:1 or SEQ ID No:2 as shown in the sequence listing.

Examples of such processes include:

- 15 a) screening a gene library which presumably contains a 029cel gene using the nucleotide sequence as a probe.
- b) preparing a primer based on the nucleotide sequence information, then performing PCR using a sample which presumably contains a 029cel gene as a template.

20 [89] More specifically, process a) above comprises:

- a) preparing a gene library which presumably contains a cellulase gene, screening the gene library using a nucleotide sequence comprising all or part of the nucleotide sequence of SEQ ID No:2 as shown in the sequence listing to select sequences which hybridize with the nucleotide sequence comprising all or part of the nucleotide sequence of SEQ ID No:2 as shown in the sequence listing from the gene library, then isolating the selected sequences, and isolating a 029cel gene from the sequences which have been selected and isolated from the gene library.
- 25

[90] The gene library may be a genomic DNA library or a cDNA library, and may be prepared according to a known procedure.

30

IV. PROTEIN EXPRESSION

[91] Proteins of the present invention are produced by culturing cells transformed with an expression vector containing the inventive cellulase gene whose expression is under control of promoter sequences. The present invention is particularly useful for enhancing the intracellular and/or extracellular production of proteins. The protein may be homologous or heterologous.

35

[92] Proteins of the present invention may also be modified in a way to form chimeric molecules comprising a protein of interest fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of the protein of interest with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxyl-terminus of the protein of interest.

[93] Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; HIS6 and metal chelation tags, the flu HA tag polypeptide and its antibody 12CA5 (Field *et al.*, *Mol. Cell. Biol.* 8:2159-2165 (1988)); the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto (Evan *et al.*, *Molecular and Cellular Biology* 5:3610-3616 (1985)); and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody (Paborsky *et al.*, *Protein Engineering* 3(6):547-553 (1990)). Other tag polypeptides include the FLAG-peptide (Hopp *et al.*, *BioTechnology* 6:1204-1210 (1988)); the KT3 epitope peptide (Martin *et al.*, *Science* 255:192-194 (1992)); tubulin epitope peptide (Skinner *et al.*, *J. Biol. Chem.* 266:15163-15166 (1991)); and the T7 gene 10 protein peptide tag (Lutz-Freyermuth *et al.*, *Proc. Natl. Acad. Sci. USA* 87:6393-6397 (1990)).

[94] Conditions appropriate for expression of said 029cel gene comprises providing to the culture the components necessary for growth and/or expression of the inventive cellulase.

Optimal conditions for the production of the proteins will vary with the choice of the host cell, and with the choice of protein to be expressed. Such conditions will be easily ascertained by one skilled in the art through routine experimentation or optimization.

[95] The protein of interest is typically purified or isolated after expression. The protein of interest may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. For example, the protein of interest may be purified using a standard anti-protein of interest antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. For general guidance in suitable purification techniques, see Scopes, *Protein Purification* (1982). The degree of purification necessary will vary depending on the use of the protein of interest. In some instances no purification will be necessary.

V. Utility of cellulase

[96] Treatment of textiles according to the present invention contemplates textile processing or cleaning with a composition comprising the cellulase of this invention. Such

treating includes, but is not limited to, stonewashing, modifying the texture, feel and/or appearance of cellulose-containing fabrics or other techniques used during manufacturing or cleaning/reconditioning of cellulose-containing fabrics. Additionally, treating within the context of this invention contemplates the removal of "immature" or "dead" cotton from cellulosic fabric or fibers. Immature cotton is significantly more amorphous than mature cotton and because of, for example, uneven dyeing. The composition contemplated in the present invention further includes a cellulase component for use in washing a soiled manufactured cellulose-containing fabric. For example, a cellulase of this invention may be used in a detergent composition for washing laundry. Detergent compositions useful in accordance with the present invention include special formulations such as pre-wash, pre-soak and home-use color restoration compositions. Such treating compositions, as described herein, may be in the form of a concentrate which requires dilution or in the form of a dilute solution or a form which can be applied directly to the cellulose-containing fabric. General treatment techniques for cellulase treatment of textiles are described in, for example, EP Publication No. 220 016 and GB Application Nos. 1,368,599 and 2,095,275.

[97] Treatment of a cellulosic material according to the present invention further contemplates the treatment of animal feed, pulp and/or paper, food and grain for purposes known in the art. For example, cellulases are known to increase the value of animal feed, improve the drainability of wood pulp, enhance food products and reduce fiber in grain during the grain wet milling process or dry milling process.

[98] Treating according to the instant invention comprises preparing an aqueous solution which contains an effective amount of a cellulase or a combination of cellulases together with other optional ingredients including, for example, a buffer, a surfactant, and/or a scouring agent. An effective amount of a cellulase enzyme composition is a concentration of cellulase enzyme sufficient for its intended purpose. Thus, for example, an "effective amount" of cellulase in a stonewashing composition according to the present invention is that amount which will provide the desired effect, e.g., to produce a worn and faded look in seams and on fabric panels. Similarly, an "effective amount" of cellulase in a composition intended for improving the feel and/or appearance of a cellulose-containing fabric is the amount that produces measurable improvements in the feel, e.g., improving the smoothness of the fabric, or appearance, e.g., removing pills and fibrils which tend to reduce the sharpness in appearance of a fabric. The amount of cellulase employed is also dependent on the equipment employed, the process parameters employed (the temperature of the cellulase treatment solution, the exposure time to the cellulase solution, and the like), and the cellulase activity (e.g., a particular solution will require a lower concentration of cellulase where a more active cellulase composition is used as compared to a less active cellulase composition). The exact concentration of cellulase in the aqueous treatment solution to

which the fabric to be treated is added can be readily determined by the skilled artisan based on the above factors as well as the desired result. In stonewashing processes, it has generally been preferred that the cellulase be present in the aqueous treating solution in a concentration of from about 0.5 to 5,000 ppm and most preferably about 10 to 200 ppm total protein. In compositions for the improvement of feel and/or appearance of a cellulose-containing fabric, it has generally been preferred that the cellulase be present in the aqueous treating solution in a concentration of from about 0.1 to 2000 ppm and most preferably about 0.5 to 200 ppm total protein.

[99] In a preferred treating embodiment, a buffer is employed in the treating composition such that the concentration of buffer is sufficient to maintain the pH of the solution within the range wherein the employed cellulase exhibits activity. The pH at which the cellulase exhibits activity depends on the nature of the cellulase employed. The exact concentration of buffer employed will depend on several factors which the skilled artisan can readily take into account. For example, in a preferred embodiment, the buffer as well as the buffer concentration are selected so as to maintain the pH of the final cellulase solution within the pH range required for optimal cellulase activity. The determination of the optimal pH range of the cellulases of the invention can be ascertained according to well-known techniques. Suitable buffers at pH within the activity range of the cellulase are also well known to those skilled in the art in the field.

[100] In addition to cellulase and a buffer, the treating composition may optionally contain a surfactant. Suitable surfactants include any surfactant compatible with the cellulase being utilized and the fabric including, for example, anionic, non-ionic and ampholytic surfactants. Suitable anionic surfactants include, but are not limited to, linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; alkanesulfonates and the like. Suitable counter ions for anionic surfactants include, but are not limited to, alkali metal ions such as sodium and potassium; alkaline earth metal ions such as calcium and magnesium; ammonium ion; and alkanolamines having 1 to 3 alkanol groups of carbon number 2 or 3. Ampholytic surfactants include, e.g., quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants generally comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, and fatty acid glycerine monoesters. Mixtures of surfactants can also be employed in manners known to those skilled in the art.

[101] A concentrated cellulase composition can be prepared for use in the methods described herein. Such concentrates contain concentrated amounts of the cellulase composition described above, buffer and surfactant, preferably in an aqueous solution.

When so formulated, the cellulase concentrate can readily be diluted with water so as to quickly and accurately prepare cellulase preparations having the requisite concentration of each constituent. When aqueous concentrates are formulated, these concentrates can be diluted so as to arrive at the requisite concentration of the components in the cellulase solution as indicated above. As is readily apparent, such cellulase concentrates permit facile formulation of the cellulase solutions as well as permit feasible transportation of the composition to the location where it will be used. The treating concentrate can be in any art-recognized form, for example, liquid, emulsion, gel, or paste. Such forms are well known to those skilled in the art.

10 [102] When a solid cellulase concentrate is employed, the cellulase composition may be a granule, a powder, an agglomerate or a solid disk. The granules can be formulated so as to contain materials to reduce the rate of dissolution of the granules into the wash medium. Such materials and granules are disclosed in U.S. Patent No. 5,254,283.

15 [103] Other materials can also be used with or placed in the cellulase composition of the present invention as desired, including stones, pumice, fillers, solvents, enzyme activators, and anti-redeposition agents depending on the eventual use of the composition.

[104] By way of example, stonewashing methods will be described in detail, however, the parameters described are readily modified by the skilled artisan for other applications, *i.e.*, improving the feel and/or appearance of a fabric. The cellulose-containing fabric is contacted with the cellulase containing stonewashing composition containing an effective amount of the cellulase by intermingling the treating composition with the stonewashing composition, and thus bringing the cellulase enzyme into proximity with the fabric. Subsequently, the aqueous solution containing the cellulase and the fabric is agitated. If the treating composition is an aqueous solution, the fabric may be directly soaked in the solution. Similarly, where the stonewashing composition is a concentrate, the concentrate is diluted into a water bath with the cellulose-containing fabric. When the stonewashing composition is in a solid form, for example a pre-wash gel or solid stick, the stonewashing composition may be contacted by directly applying the composition to the fabric or to the wash liquor.

30 [105] The cellulose-containing fabric is incubated with the stonewashing solution under conditions effective to allow the enzymatic action to confer a stonewashed appearance to the cellulose-containing fabric. For example, during stonewashing, the pH, liquor ratio, temperature and reaction time may be adjusted to optimize the conditions under which the stonewashing composition acts. "Effective conditions" necessarily refers to the pH, liquor ratio, and temperature which allow the cellulase enzyme to react efficiently with cellulose-containing fabric, in this case to produce the stonewashed effect. It is within the skill of

those in the art to maximize conditions for using the stonewashing compositions according to the present invention.

[106] The liquor ratios during stonewashing, *i.e.*, the ratio of weight of stonewashing composition solution (*i.e.*, the wash liquor) to the weight of fabric, employed herein is

generally an amount sufficient to achieve the desired stonewashing effect in the denim fabric and is dependent upon the process used. Preferably, the liquor ratios are from about 4:1 to about 50:1; more preferably from about 5:1 to about 20:1, and most preferably from about 10:1 to about 15:1.

[107] Reaction temperatures during stonewashing with the present stonewashing compositions are governed by two competing factors. Firstly, higher temperatures generally correspond to enhanced reaction kinetics, *i.e.*, faster reactions, which permit reduced reaction times as compared to reaction times required at lower temperatures. Accordingly, reaction temperatures are generally at least about 10°C and greater. Secondly, cellulase is a protein which loses activity beyond a given reaction temperature, which temperature is dependent on the nature of the cellulase used. Thus, if the reaction temperature is permitted to go too high, the cellulolytic activity is lost as a result of the denaturing of the cellulase. While standard temperatures for cellulase usage in the art are generally in the range of 35°C to 65°C, and these conditions would also be expected to be suitable for the cellulase of the invention, the optimal temperature conditions should be ascertained according to well known techniques with respect to the specific cellulase used.

[108] Reaction times are dependent on the specific conditions under which the stonewashing occurs. For example, pH, temperature and concentration of cellulase will all affect the optimal reaction time. Generally, reaction times are from about 5 minutes to about 5 hours, and preferably from about 10 minutes to about 3 hours and, more preferably, from about 20 minutes to about 1 hour.

[109] According to yet another preferred embodiment of the present invention, the cellulase of the invention may be employed in a detergent composition. The detergent compositions according to the present invention are useful as pre-wash compositions, pre-soak compositions, or for cleaning during the regular wash or rinse cycle. Preferably, the detergent composition of the present invention comprises an effective amount of cellulase, a surfactant, and optionally includes other ingredients described below.

[110] An effective amount of cellulase employed in the detergent compositions of this invention is an amount sufficient to impart the desirable effects known to be produced by cellulase on cellulose-containing fabrics, for example, depilling, softening, anti-pilling, surface fiber removal, anti-graying and cleaning. Preferably, the cellulase in the detergent composition is employed in a concentration of from about 10 ppm to about 20,000 ppm of detergent.

[111] The concentration of cellulase enzyme employed in the detergent composition is preferably selected so that upon dilution into a wash medium, the concentration of cellulase enzyme is in a range of about 0.01 to about 1000 ppm, preferably from about 0.02 ppm to about 500 ppm, and most preferably from about 0.5 ppm to about 250 ppm total protein.

5 The amount of cellulase enzyme employed in the detergent composition will depend on the extent to which the detergent will be diluted upon addition to water so as to form a wash solution.

[112] The detergent compositions of the present invention may be in any art recognized form, for example, as a liquid, in granules, in emulsions, in gels, or in pastes. Such forms
10 are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase is preferably formulated as granules. Preferably, the granules can be formulated so as to additionally contain a cellulase protecting agent. The granule can be formulated so as to contain materials to reduce the rate of dissolution of the granule into the wash medium. Such materials and granules are disclosed in U.S. Patent No. 5,254,283.

15

[113] The detergent compositions of this invention employ a surface active agent, *i.e.*, surfactant, including anionic, non-ionic and ampholytic surfactants well known for their use in detergent compositions.

[114] Suitable anionic surfactants for use in the detergent composition of this invention
20 include linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; and alkanesulfonates. Suitable counter ions for anionic surfactants include alkali metal ions such as sodium and potassium; alkaline earth metal ions such as calcium and magnesium; ammonium ion; and alkanolamines having 1 to 3 alkanol groups of carbon number 2 or 3.
25 Ampholytic surfactants include quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants generally comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like. Suitable surfactants for use in this invention
30 are disclosed in British Patent Application No. 2 094 826 A.

Mixtures of such surfactants can also be used. The surfactant or a mixture of surfactants is generally employed in the detergent compositions of this invention in an amount from about 1 weight percent to about 95 weight percent of the total detergent composition and preferably from about 5 weight percent to about 45 weight
35 percent of the total detergent composition. In addition to the cellulase composition and the surfactant(s), the detergent compositions of this invention can optionally contain one or more of the following components:

Hydrolases Except Cellulase

[115] Suitable hydrolases include carboxylate ester hydrolase, thioester hydrolase, phosphate monoester hydrolase, and phosphate diester hydrolase which act on the ester bond; glycoside hydrolase which acts on glycosyl compounds; an enzyme that hydrolyzes N-glycosyl compounds; thioether hydrolase which acts on the ether bond; and α -amino-acyl-peptide hydrolase, peptidyl-amino acid hydrolase, acyl-amino acid hydrolase, dipeptide hydrolase, and peptidyl-peptide hydrolase which act on the peptide bond. Preferable among them are carboxylate ester hydrolase, glycoside hydrolase, and peptidyl-peptide hydrolase. Suitable hydrolases include (1) proteases belonging to peptidyl-peptide hydrolase such as pepsin, pepsin B, rennin, trypsin, chymotrypsin A, chymotrypsin B, elastase, enterokinase, cathepsin C, papain, chymopapain, ficin, thrombin, fibrinolysin, renin, subtilisin, aspergillopeptidase A, collagenase, clostridiopeptidase B, kallikrein, gastrisin, cathepsin D., bromelin, keratinase, chymotrypsin C, pepsin C, aspergillopeptidase B, urokinase, carboxypeptidase A and B, and aminopeptidase; (2) glycoside hydrolases (cellulase which is an essential ingredient is excluded from this group) α -amylase, β -amylase, gluco amylase, invertase, lysozyme, pectinase, chitinase, and dextranase. Preferably among them are α -amylase and β -amylase. They function in acid to neutral systems, but one which is obtained from bacteria exhibits high activity in an alkaline system; (3) carboxylate ester hydrolase including carboxyl esterase, lipase, pectin esterase, and chlorophyllase. Especially effective among them is lipase.

[116] The hydrolase other than cellulase is incorporated into the detergent composition as much as required according to the purpose. It should preferably be incorporated in an amount of 0.001 to 5 weight percent, and more preferably 0.02 to 3 weight percent, in terms of purified protein. This enzyme should be used in the form of granules made of crude enzyme alone or in combination with other components in the detergent composition. Granules of crude enzyme are used in such an amount that the purified enzyme is 0.001 to 50 weight percent in the granules. The granules are used in an amount of 0.002 to 20 and preferably 0.1 to 10 weight percent. As with cellulases, these granules can be formulated so as to contain an enzyme protecting agent and a dissolution retardant material.

Cationic Surfactants and Long-Chain Fatty Acid Salts

[117] Such cationic surfactants and long-chain fatty acid salts include saturated or unsaturated fatty acid salts, alkyl or alkenyl ether carboxylic acid salts, α -sulfofatty acid salts or esters, amino acid-type surfactants, phosphate ester surfactants, quaternary ammonium salts including those having 3 to 4 alkyl substituents and up to 1 phenyl substituted alkyl substituents. Suitable cationic surfactants and long-chain fatty acid salts are disclosed in British Patent Application No. 2 094 826 A.

The composition may contain from about 1 to about 20 weight percent of such cationic surfactants and long-chain fatty acid salts.

Builders

A. Divalent sequestering agents

5 [118] The composition may contain from about 0 to about 50 weight percent of one or more builder components selected from the group consisting of alkali metal salts and alkanolamine salts of the following compounds: phosphates, phosphonates, phosphonocarboxylates, salts of amino acids, aminopolyacetates high molecular electrolytes, non-dissociating polymers, salts of dicarboxylic acids, and aluminosilicate salts. Suitable divalent sequestering agents
10 are disclosed in British Patent Application No. 2 094 826 A.

B. Alkalis or inorganic electrolytes

[119] The composition may contain from about 1 to about 50 weight percent, preferably from about 5 to about 30 weight percent, based on the composition of one or more alkali
15 metal salts of the following compounds as the alkalis or inorganic electrolytes: silicates, carbonates and sulfates as well as organic alkalis such as triethanolamine, diethanolamine, monoethanolamine and triisopropanolamine.

Antiredeposition Agents

[120] The composition may contain from about 0.1 to about 5 weight percent of one or
20 more of the following compounds as antiredeposition agents: polyethylene glycol, polyvinyl alcohol, polyvinylpyrrolidone and carboxymethylcellulose.

[121] Among them, a combination of carboxymethyl-cellulose and/or polyethylene glycol with the cellulase composition of the present invention provides for an especially useful dirt removing composition.

Bleaching Agents

[122] The use of the cellulase of the present invention in combination with a bleaching agent such as potassium monopersulfate, sodium percarbonate, sodium perborate, sodium sulfate/hydrogen peroxide adduct and sodium chloride/hydrogen peroxide adduct or/and a photo-sensitive bleaching dye such as zinc or aluminum salt of sulfonated phthalocyanine
30 further improves the detergenting effects. Similarly, bleaching agents and bleach catalysts as described in EP 684 304 may be used.

Bluing Agents and Fluorescent Dyes

[123] Various bluing agents and fluorescent dyes may be incorporated in the composition, if necessary. Suitable bluing agents and fluorescent dyes are disclosed in British Patent
35 Application No. 2 094 826 A.

Caking Inhibitors

[124] The following caking inhibitors may be incorporated in the powdery detergent: p-toluenesulfonic acid salts, xylenesulfonic acid salts, acetic acid salts, sulfosuccinic acid salts, talc, finely pulverized silica, amorphous silicas, clay, calcium silicate (such as Micro-Cell™ of Johns Manville Co.), calcium carbonate and magnesium oxide.

5 Antioxidants

[125] The antioxidants include, for example, tert-butyl-hydroxytoluene, 4,4'-butylidenebis(6-tert-butyl-3-methylphenol), 2,2'-butylidenebis(6-tert-butyl-4-methylphenol), monostyrenated cresol, distyrenated cresol, monostyrenated phenol, distyrenated phenol and 1,1-bis(4-hydroxy-phenyl)cyclohexane.

10 Solubilizers

[126] The solubilizers include, for example, lower alcohols such as ethanol, benzenesulfonate salts, lower alkylbenzenesulfonate salts such as p-toluenesulfonate salts, glycols such as propylene glycol, acetylbenzene-sulfonate salts, acetamides, pyridinedicarboxylic acid amides, benzoate salts and urea.

15 [127] The detergent composition of the present invention can be used in a broad pH range from acidic to alkaline pH. In a preferred embodiment, the detergent composition of the present invention can be used in mildly acidic, neutral or alkaline detergent wash media having a pH of from above 5 to no more than about 12.

20 [128] Aside from the above ingredients, perfumes, buffers, preservatives, dyes and the like can be used, if desired, with the detergent compositions of this invention. Such components are conventionally employed in amounts heretofore used in the art.

25 [129] When a detergent base used in the present invention is in the form of a powder, it may be one which is prepared by any known preparation methods including a spray-drying method and a granulation method. The detergent base obtained particularly by the spray-drying method, agglomeration method, dry mixing method or non-tower route methods are preferred. The detergent base obtained by the spray-drying method is not restricted with respect to preparation conditions. The detergent base obtained by the spray-drying method is hollow granules which are obtained by spraying an aqueous slurry of heat-resistant ingredients, such as surface active agents and builders, into a hot space. After the spray-drying, perfumes, enzymes, bleaching agents, inorganic alkaline builders may be added.
30 With a highly dense, granular detergent base obtained such as by the spray-drying-granulation or agglomeration method, various ingredients may also be added after the preparation of the base.

35 [130] When the detergent base is a liquid, it may be either a homogeneous solution or a nonhomogeneous dispersion. For removing the decomposition of carboxymethylcellulose by

the cellulase in the detergent, it is desirable that carboxymethylcellulose is granulated or coated before the incorporation in the composition.

[131] The detergent compositions of this invention may be incubated with cellulose-containing fabric, for example soiled fabrics, in industrial and household uses at temperatures, reaction times and liquor ratios conventionally employed in these environments.

[132] Detergents according to the present invention may additionally be formulated as a pre-wash in the appropriate solution at an intermediate pH where sufficient activity exists to provide desired improvements softening, depilling, pilling prevention, surface fiber removal or cleaning. When the detergent composition is a pre-soak (e.g., pre-wash or pre-treatment) composition, either as a liquid, spray, gel or paste composition, the cellulase enzyme is generally employed from about 0.0001 to about 1 weight percent based on the total weight of the pre-soak or pre-treatment composition. In such compositions, a surfactant may optionally be employed and when employed, is generally present at a concentration of from about 0.005 to about 20 weight percent based on the total weight of the pre-soak. The remainder of the composition comprises conventional components used in the pre-soak, *i.e.*, diluent, buffers, other enzymes (proteases), and the like at their conventional concentrations.

[133] It is contemplated that compositions comprising cellulase enzymes described herein can be used in home use as a stand alone composition suitable for restoring color to faded fabrics (see, for example, U.S. Patent No. 4,738,682)

as well as used in a spot-remover and for depilling and anti-pilling (pilling prevention).

[134] The use of the cellulase according to the invention may be particularly effective in feed additives and in the processing of pulp and paper. These additional industrial applications are described in, for example, PCT Publication No. 95/16360 and Finnish Granted Patent No. 87372, respectively.

[135] In order to further illustrate the present invention and advantages thereof, the following specific examples are given with the understanding that they are being offered to illustrate the present invention and should not be construed in any way as limiting its scope.

EXAMPLES

[136] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1 Sample collection and processing

[137] This example illustrates how to collect samples and process them to obtain sufficient DNA to create a cDNA library.

[138] Samples of water (250 ml) were collected from the littoral zone of Sonachi (Crater) Lake, Kenya using a 250-ml stainless steel beaker mounted on the end of a flexible extendible 1-m pole and placed in sealable plastic containers (Whirlpak) for transport to the laboratory at ambient temperature. The temperature of the surface waters was 28°C, with pH 10 and a conductivity of 7.23 mS cm⁻¹ (at 27°C).

[139] To collect the microbial flora, water (750 ml) from Sonachi (Crater) Lake, Kenya was filtered on site (using a hand operated vacuum pump) through a sequence of sterile membrane filters (47 mm diameter), composed of cellulose nitrate or cellulose acetate, of decreasing pore size, until all water flow stopped. The sequence of filters was 8µm, 3µm and 0.22µm. The individual membrane filters were placed immediately into 10 ml of cold, sterile cell stabilization buffer (TES) containing 10 mM Tris HCl, pH8.0; 1 mM EDTA and 5% w/v NaCl in 30 ml sterile plastic universal tubes and kept on ice in a refrigerated cool box until they could be processed further, usually within 4 hours of sampling. The microbial material on the filters was dispersed by vigorous vortex mixing with sterile glass beads (5 ml) and the cells pelleted in microfuge tubes by centrifugation at 13,000g for 5 min. The microbial material was aliquoted to the microfuge tubes in volumes estimated to contain the equivalent of 10⁸ to 10⁹ bacterial cells, giving a total of 12 tubes. The DNA was extracted using the GenomicPrep™ Cells and Tissue DNA isolation kit (Amersham Pharmacia Biotech, Piscataway, NJ, USA) following the manufacturer's instructions. Cells in each tube were resuspended in 600 µl of the Cell Lysis Solution provided, and incubated at 80°C for 5 min to lyse the cells. Samples prepared by this method are stable at room temperature for at least 18 months, and were transported back to the laboratory in this form. DNA extraction was completed by RNase A treatment, protein precipitation and isopropanol precipitation of the DNA following the manufacturer's protocol. Each DNA pellet was dissolved in 100µl sterile Tris buffer 10mM pH 8.5.

[140] DNA yield was estimated by running 5µl samples on a 0.5% w/v agarose gel and comparing with known amounts of bacterial genomic DNA. The samples were pooled, giving a total of about 20µg DNA. Since yields were low, the material was supplemented with about 30% extra material extracted from the water samples which were collected at the same time as the on-site material and stored at 4°C in the laboratory until required. This amount of DNA, about 30µg, was the amount of starting material that preliminary experiments had shown was needed to carry out the trial and bulk restriction digestion and size fractionation to give sufficient material for library construction.

Example 2

Library Construction

[141] The following example details how to prepare a DNA library for use in screening and detection of novel sequences in *E. coli*.

5 Preparation of DNA

[142] The pooled DNA was used for construction of the genomic DNA library. The purified DNA was partially digested with *Sau3A1* to give an average fragment size of about 5 kb. Restricted DNA was size fractionated by electrophoresis on 0.5% agarose in TAE (0.04M Tris-acetate, 0.001M EDTA pH 8.0). Material in the 1.5 to 10 kb range was excised and
10 replaced in a well of the same size cut in an unused part of the agarose gel and concentrated to a narrow band by reversed electrical current. The DNA band was excised and DNA extracted using the QIAGEN (Crawley, UK) QIAEXII gel extraction kit, following the manufacturer's guidelines. The eluted DNA was precipitated with ethanol and resuspended in 10 mM Tris HCl buffer, pH 8.5.

15 Preparation of Lambda libraries

[143] The restricted DNA was cloned into a Lambda vector using the ZAP-Express™ vector kit (predigested with *Bam*H1 and alkaline phosphatase treated) and the Gigapak® III Gold packaging extract (Stratagene, Amsterdam, The Netherlands) following the manufacturer's protocol. The primary libraries were amplified as per protocol by plating
20 aliquots containing $\sim 5 \times 10^4$ pfu with host *E. coli* strain XL1-Blue MRF' on 150mm Petri dishes and eluting the phage in buffer. Amplified libraries were stored in 7% v/v dimethyl sulphoxide at -80°C after freezing in liquid nitrogen. The total primary titre was 1.8×10^6 pfu and after amplification 6.8×10^9 pfu ml⁻¹.

25 Assessment of Library Quality

[144] The phagemid vector pBK-CMV was excised from the Lambda ZAP library using ExAssist helper phage (Stratagene) as described by the manufacturer, and used to infect *E. coli* strain XL0LR. Plasmid-containing clones were isolated by plating on Luria – Bertani (LB) agar containing 50 µg ml⁻¹ kanamycin. Blue:white screening in the presence of Xgal [5-
30 bromo-4-chloro-3-indoyl-β-D-galactoside] and IPTG [isopropylthio-β-D-galactoside] was used to determine cloning efficiency. If no DNA has been cloned into the Lambda vector, the β-galactosidase gene is expressed in the presence of the inducer IPTG, resulting in cleavage of the substrate analogue Xgal to produce a blue pigment in the colony. If however a fragment of the genomic DNA has been successfully cloned into the Lambda vector it
35 disrupts the gene so that no enzyme is produced and the colony remains white. The ratio of blue to white colonies therefore can be used to calculate the percentage of clones containing an insert. For this library the blue:white screen gave a ratio of 7 blue to 286 white colonies,

indicating that 97% of the clones contained an insert of the genomic DNA. Twenty four colonies were selected at random and plasmid DNA prepared using the Wizard®Plus SV Miniprep DNA purification system (Promega UK, Southampton) Restriction analysis using *Pst*I and *Hind*III which flank the *Bam*H1 cloning site followed by agarose gel electrophoresis was used to determine insert sizes. One clone out of the 24 was found to have no detectable insert. The rest had inserts ranging from 1.5 kb to 8.0 kb.

Example 3 Library screening for cellulases

[145] DNA libraries in the pBK-CMV phagemid were screened for cellulase activity in a plate assay of the *E. coli* clones. To detect cellulase activity the genomic libraries were plated on LB agar containing kanamycin, 0.5% w/v carboxymethylcellulose (low viscosity sodium salt; Sigma, Poole, UK) and IPTG (15 µl of a 0.5 M solution spread on the surface of the agar in a 7 cm diameter Petri dish). Following overnight growth at 37°C, the colonies were overlaid with 3ml molten 0.7% w/v agarose dissolved in water which had been cooled to 50°C. After this had set, the plates were flooded with 0.1% w/v Congo Red solution for 30 minutes followed by 2 washes with 1 M NaCl. Positive clones exhibiting extracellular cellulase activity were surrounded by a yellow halo against a red background (R. Teather and P.J. Wood, Applied & Environmental Microbiology, 43: 770-780, 1982).

[146] The screening of 110,000 *E.coli* pBK-CMV clones yielded 4 zones of clearing indicating potential cellulase-producing colonies. Three of these were successfully recovered as cellulase-producing clones after homogenising the agar plug removed from the cleared zone, streaking out for single colonies and confirming the phenotype by the Congo Red test.

Example 4 Characterisation of a cellulase-positive clone

[147] Plasmid DNA was isolated from the three cellulase positive clones, and the size of the inserts determined by restriction digestion as described above. All three had the same size (about 3.5kb) and the same size fragments after digestion as determined by gel electrophoresis. This indicated that all three isolates were identical, derived by amplification of a single clone. This was confirmed by the first round of sequencing of the plasmid DNA (using primer sites in the pBKCMV plasmid). This was carried out by the Protein and Nucleic Acid Chemistry Laboratory at Leicester University, using the Perkin Elmer 'BigDye' terminator chemistry and the model 377 ABI automated DNA sequencer. Complete coverage of the sequence was obtained by 'primer walking' from both the 5' and 3' ends of the insert. The sequence was edited using Applied Biosystems multisequence editor Seqed™ version 1.0.3. Sequence was assembled with programmes in the GCG Wisconsin

Package, version 10.2-UNIX, available at the University of Leicester. This identified an insert of environmental DNA of 3410 nucleotide bases (Figure 1).

Example 5 Identification of the cellulase-gene

5 [148] Possible Open Reading Frames (ORF) in the nucleotide sequence of the inserted environmental DNA of clone 029cel were identified using the ORF Find facility of the MapDraw program (DNASTAR, Brighton, MA, USA) or ORF Search from the Vector NTI Suite of programs (InforMax®, North Bethesda, MD, USA).

10 [149] This identified an ORF composed of 1746 nucleotides corresponding to a protein of 581 amino acids, starting at position 3004 of the insert sequence and ending at position 1259. The sequence of this ORF was excised using EditSeq (DNASTAR) and examined by BLAST programs.

[150] The nucleotide sequence of this ORF is shown in Figure 2.

15 [151] An examination of the nucleotide sequence using the BLASTn program, which compares a nucleotide query sequence against a non-redundant nucleotide sequence database, indicated no significant similarity to any known sequence.

[152] An examination of the nucleotide sequence using the BLASTx program, which compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database, revealed surprisingly, very low similarity
20 (no more than 29%) to a number of bacterial endocellulases. The highest alignment score revealed 25% identity (132 amino acids) to a 527 amino acid region of CelJ, an enzyme comprising 1601 amino acids, the largest catalytic component of the cellulosome of *Clostridium thermocellum* (protein id BAA1207070.1, accession D83704.1). It is very probable that an enzyme with homology this low would not have been detected using
25 convention methods using DNA probes based on known cellulase gene sequences, especially given the very high diversity of cellulases already characterised.

[153] The translated protein composed of 581 amino acids is shown in Figure 3.

Example 6 Enzyme Characterization

30 Influence of salt

[154] Cells of *E.coli* pBK-CMV containing the 029cel gene are suspended in 5 ml buffer (20 mM TRIS-HCl, pH8.0; 500 mM NaCl; 0.1 mM EDTA; 0.1% Triton X-100™) and disrupted by sonication on ice. The sonicated extracts are examined by agar diffusion assay on carboxymethylcellulose (CMC) at different NaCl concentrations. Sonicated extracts (100 µL)
35 and 1 in 10 dilutions are placed in wells punched in CMC-agar plates containing varying amounts of NaCl. The plates are incubated at 37°C for 16 hours and the resulting clearing

zones indicating cellulose hydrolysis measured in millimetres. The cellulase 029cel is active over the range 0 – 25% w/v NaCl, although the activity at 25% w/v NaCl is lower than the activity at 0% NaCl.

5 Influence of pH

[155] The influence of pH on cellulase activity is investigated using the pH-gradient plate method described by Grant & Tindall (Isolation of alkaliphilic bacteria, In: Microbial Growth and Survival in Extreme Environments, Academic Press, London, 1980, pp. 27-36). An agar medium containing CMC is poured to a depth of 1 cm in square Petri dishes and allowed to
10 set. A uniform trough 1 cm wide is cut from one edge of the plate and agar containing 20% w/v $\text{Na}_2\text{CO}_3 \cdot 10\text{H}_2\text{O}$ and 0.2 M NaOH (prepared by mixing equal volumes of sterile 0.4 M NaOH/40% w/v $\text{Na}_2\text{CO}_3 \cdot 10\text{H}_2\text{O}$ and 4% w/v agar at 60°C) is poured into the trough. The plates are developed at 37°C overnight to allow a uniform gradient from pH 12 to pH 7 to form. To test the pH tolerance of the 029cel cellulase a narrow trough is cut through the
15 (agar) gradient at right angles to the original trough and filled with 1 ml of sonicated cell extract. The plates were allowed to develop overnight at 37°C. The plates are treated with Congo Red for 30 minutes to visualize the zone of cellulose hydrolysis. The results indicate that the 029cel cellulase is active to about pH 11.5.

20

[156] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

25

SEQUENCE LISTING

<110> Genencor International, Inc.

<120> Novel Bacillus 029cel Cellulase

<130> 11816-107

<140> CA 2,523,328

<141> 2004-04-28

<150> US 60/466,831

<151> 2003-04-29

<160> 3

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 3410

<212> DNA

<213> Bacillus sp.

<220>

<221> misc_feature

<222> (1)...(3410)

<223> isolated from environmental sample from Sonachi Lake, Kenya

<400> 1

atcaacacgc	tggaaagtaa	tttcaagggt	aaggccatcg	gttgccgccg	gggtagaaat	60
gtgcggttg	atttcgttga	gcggcgtcgc	cggcgttcca	ccgagggcat	agcgcagcag	120
gttggcgatg	ccaccggtga	ggccttcggg	gccgcctacg	atgttgtgct	cagccgcccc	180
tgcgatgtag	ccgtccggct	cgggttcgct	cgcgggggtg	agaagacaa	tgtcgtcgag	240
ataaaggttg	ccgcttccgc	tctcaacgcc	gccgaggttg	aattggattt	cgcaaattct	300
cgttaggtcc	agcacggaat	cgccgacgag	gtcggctatg	ggaatctgaa	tgcgcccata	360
gggttgggta	cgcggaaggg	acacgtaggg	acccactttg	tcattgggcg	agacgagccg	420
gacaaagatt	tgggtgcgccg	cctgcgaggg	gccttgaggg	gcgagagaaa	ggtacgtgag	480
ggcgctgatg	tcgtgcgtgg	gaccgtctcc	ccagttgtcg	agattgagcc	caaatccggc	540
ccaccatccg	gcgatagtgt	agctccaatg	gtagtgaacg	tcaccctcga	agccgcccgt	600
ggagagttcc	tgcaagccgt	cgccccaaat	gcccgtgatg	agcgttgcc	cgtcacggta	660
gatcacaagt	tcggcgggcg	gtgcccgggg	aagatcgcc	tgagtgatca	cgagagtggc	720
ggtggcgctg	ccttcgtgat	tagggctcgt	aatggtggcg	acgaccgtgt	agctaccggg	780
ccccactggc	gcatgggtgg	aaccgttgta	ggtaaaggag	acgtcaagcc	ccacgggatg	840
ggtctcggca	agagcggcct	tgggggtgcc	gtcgaaaacg	tgttccaaat	tggagagcgt	900
gatggtggcg	ggtgccttga	gcacagtcac	agaaacagtg	gattgcacgg	gatcgtgcgc	960
tgccgtgtct	gcaggtgtga	agaccacgct	gtaaaaacgg	gttcggcgcg	acggtgcaag	1020
gccggacagg	acaaaggcaa	agtcgccggg	gacggcggtc	actccgccgc	tcaggccggc	1080
ctccgcaagg	gtttgcccga	aggtgatggg	tgcggctgtg	ggccacatct	ccacaaggcc	1140
ggtgtcccc	tcgtcacgca	ccggcatgag	ggcgagagag	agatgaatgt	aactggcttg	1200
gtaattgatg	tcgggctcgg	tgatttccca	tgagttctcc	ggccaaaaac	cattccaatc	1260
aaggtaggct	ttttgcacgg	gttggtctcg	gatcgccctga	atgcttccgc	tgtatttggg	1320
cattgggacc	cgcccgaag	aaaaccagga	gcgggaccgt	agagtgaagt	gagggcattg	1380
tcccagtcg	gccatcgcg	aaccaatggt	ggtagatttc	attggctgca	cggtcagcgc	1440
cgctggcata	catgttgcta	agatagacca	tgcccattgg	gttcaactcc	tggagatagt	1500
gcaggtagcc	catcgcgcca	tcgcgatgcg	cggccgcgtc	ggcggggttg	agcccaagcc	1560
tccgtacccc	ctcgaagaaa	aagccagcct	gagactttgt	tttggttcgag	ccccacgtgt	1620
aatcctgatc	cttcaggtag	gcgcggtagg	cgtcggctctg	gttattccat	gcaccgagaa	1680
actccccacc	gtttatagaa	gccgccatcc	ggttgccgat	gtcggcagag	acgctaggcg	1740
tcgctcccgg	gagggctcgtg	tagtgggcga	gagctttttg	tagctcacct	tgaaagggga	1800
agaaatacca	ccactgcacg	ggctccatat	cgagatagcg	cacatcgaag	aaatcgcgat	1860
agaccgcacc	gcccgtgcgc	tcgaagagca	tggcgggcgg	catcacacgg	ttggctagcg	1920

tatcgtgggc	attgcgcgag	gggctcacgg	aagcaaatcc	gggtgtgtcg	aaaggcacat	1980
gaggatggac	catgggtccaa	ttccatgcgg	cgatggcagc	ggattcgagg	gtgacggcat	2040
aatcgctcat	gcctacgctc	tcaaagacag	tcgccccgag	ggcgaaagcg	gcggcagcca	2100
tggcagtggc	ctcggtcgag	acggggccgt	agtaacgcgg	atgggtgtcg	gtgctcggcg	2160
ggctggcgct	ctgggtgcccc	gtcacggaaa	ctttcccagag	aatagccccg	ctcggctcct	2220
gcatgcgtaa	gagccagtcc	attccccatt	tgacttcgtc	aagcaggctc	gggacaccgt	2280
tgccggattc	cgggatgcc	aaatcatcgg	taaagacgtc	aggccgccct	tgataggcaa	2340
ggagcagctc	caggatgacg	cgccccgtcc	actcgtgtga	cttggtgaaa	tcgcccgcac	2400
cgaaccaacc	gccgctgaga	tcgcgctcca	aggaggcatt	ccccatatcc	cagatggggc	2460
ggctggcgac	gtcctgcggg	tgagaagcgg	catcggtcca	gttcgcgtgg	gcgtagggca	2520
cctccttggc	aaacccggag	cgctgataga	agaacatgcg	cacggcctcg	cgcaggacaa	2580
catcgtaaac	atccgcgcca	atggcgaaac	tatcggaatg	agtgtgtgtg	gcaggatcgt	2640
ggatgcggta	gtggccgggg	tcggcaacta	ccgtaaaatc	aaaccaccac	acgcggtctc	2700
ccgattgaat	atggatggcg	ccgccgttcc	acgggaccgg	tgagccggag	aaaaccacga	2760
cgccatcggt	cacgcgacgg	acctccagcg	ttgcgcgggg	gctgtagctc	tcggcgctgt	2820
tccagccaat	ctgcgggtcg	gcgatcaccg	ccaccttggt	ggcatcgggc	gggtaaccga	2880
attggtcgat	gcggatttta	tcgggtgtgg	tggaggcgac	gagggcgagg	ctgcccataga	2940
gcagcaagaa	aaagcccgtc	gtcggcccga	tacaaaaaaa	acgaataggg	agagaaaaat	3000
tcatagcagg	atgtggatac	ggaaaggggg	aaaacggtgc	aaagacccaa	gcccacgcgt	3060
tggcgaaaac	tggatggttg	gtttatcaag	aaaagcgctt	ttgagccaaa	agctgcgggc	3120
aatccttatt	gcgtttcaca	atattttcac	atcgtcggcg	gcacgacttt	tcgatgggcg	3180
acttgacagc	gtattctctc	aggcgcgagg	ctgcaaacct	tatgaaaaaa	ggcccgcgca	3240
gcgatctgtc	cccgttcaaa	atccagtcaa	ggtttgttca	agggtttgag	gtctgataga	3300
ggcacagtcg	agccatcagc	agtcgcattg	agtaggggtg	ttggagaaa	gtgtgcaaatg	3360
accgctgccg	aaggaactgt	ggagacaaaa	agcatatttt	cctcgccaag		3410

<210> 2

<211> 1746

<212> DNA

<213> Bacillus sp.

<220>

<221> misc_feature

<222> (1)...(1746)

<223> isolated from environmental sample from Sonachi Lake, Kenya

<400> 2

atgaattttt	ctctccctat	tcgttttttt	ggtatcgggc	cgacagcggg	ctttttcttg	60
ctgctcatgg	gcagctccgc	cctcgtcgcc	tccaccacaa	ccgataaaa	ccgcatcgac	120
caattcgggt	accccgccga	tgccaccaag	gtggcggtga	tcgccgaccc	gcagattggc	180
tggaaacagc	ccgagagcta	cagccccggc	gcaacgctgg	aggtccgtcg	cgtgaacgat	240
ggcgctcgtg	ttttctccgg	ctcaccggtc	ccgtggaacg	gcggcgccat	ccatattcaa	300
tcgggagacc	gcgtgtggtg	gtttgatatt	acggtagttg	ccgagcccgg	ccactaccgc	360
atccacgatc	ctgccaaaca	cactcattcc	gatagtttcg	ccattggcgc	ggatgtttac	420
gatgtttgtc	tgcgcgaggc	cgtgcgcatg	ttcttctatc	agcgctccgg	gtttgccaag	480
gaggtgccct	acgcccacgc	gaactgggcc	gatgccgctt	ctcaccgcga	ggacgtcgcc	540
agccgcccc	tctgggatat	ggggaatgcc	tccttgagac	gcgatctcag	cggcggttgg	600
ttcgatgcgg	gcgatttcaa	caagtacagc	gagtggacgg	ggcgcgctcat	cctggagctg	660
ctccttgcc	atcaaggggc	gcctgacgtc	tttaccgatg	attttgcat	cccggaatcc	720
ggcaacgggt	tccccgacct	gcttgacgaa	gtcaaatggg	gaatggactg	gctcttacgc	780
atgcaggagc	cgagcggggc	tattctcggg	aaagtttccg	tgacggggca	ccagagcgcc	840
agcccgccga	gcaccgacac	ccatccgcgt	tactacggcc	ccgtctcgac	cgaggccact	900
gccatggctg	ccgccgcttt	cgccctcggg	gcgactgtct	ttgagagcgt	aggcatgagc	960
gattatgccg	tcaccctcga	atccgctgcc	atcgccgcat	ggaattggac	catggtccat	1020
cctcatgtgc	ctttcgacaa	caccggattt	gcttccgtga	gccccctcgc	caatgcccac	1080
gatacgctag	ccaaccgtgt	gatggccgcc	gccatgctct	tcgagcgcac	ggcggtgctg	1140
gtctatcgcg	atcttctcga	tgtgcgctat	ctcgatatgg	agcccggtga	gtggtggtat	1200
ttcttcccct	ttcaagggtga	gctacaaaaa	gctctcgccc	actacacgac	cctcccggga	1260
gcgacgccta	gcgtctctgc	cgacatccgc	aaccggatgg	cggcttctat	aaacgggtggg	1320
gagtttctcg	gtgcatggaa	taaccagacc	gacgcctacc	gcgcctacct	gaaggatcag	1380
gattacacgt	ggggctcgaa	caaaacaaag	tctcaggctg	gctttttctt	cgagggggta	1440

```

cggaggcttg ggctcaaccc cgccgacgcg gccgcgcac gcgatgccgc gatgggctac 1500
ctgcactatc tccacggagt gaacccaatg ggcatggtct atcttagcaa catgtatgcc 1560
agcggcgctg accgtgcagc caatgaaatc taccaccatt ggttccgcga tggccggact 1620
gggacaatgc cctcacttca ctctacggtc ccgctcctgg ttttctttcg ggcgggtccc 1680
aatgccc aaa tacagcggaa gcattcaggc gatccgagac caaccctgc aaaaagccta 1740
ccttga 1746

```

<210> 3

<211> 581

<212> PRT

<213> Bacillus sp.

<220>

<221> VARIANT

<222> (1)...(581)

<223> isolated from environmental sample from Sonachi Lake, Kenya

<400> 3

```

Met Asn Phe Ser Leu Pro Ile Arg Phe Phe Gly Ile Gly Pro Thr Ala
 1          5          10          15
Gly Phe Phe Leu Leu Leu Met Gly Ser Ser Ala Leu Val Ala Ser Thr
      20          25          30
His Thr Asp Lys Ile Arg Ile Asp Gln Phe Gly Tyr Pro Ala Asp Ala
      35          40          45
Thr Lys Val Ala Val Ile Ala Asp Pro Gln Ile Gly Trp Asn Ser Ala
      50          55          60
Glu Ser Tyr Ser Pro Gly Ala Thr Leu Glu Val Arg Arg Val Asn Asp
65          70          75          80
Gly Val Val Val Phe Ser Gly Ser Pro Val Pro Trp Asn Gly Gly Ala
      85          90          95
Ile His Ile Gln Ser Gly Asp Arg Val Trp Trp Phe Asp Phe Thr Val
      100          105          110
Val Ala Glu Pro Gly His Tyr Arg Ile His Asp Pro Ala Asn Asn Thr
      115          120          125
His Ser Asp Ser Phe Ala Ile Gly Ala Asp Val Tyr Asp Val Val Leu
      130          135          140
Arg Glu Ala Val Arg Met Phe Phe Tyr Gln Arg Ser Gly Phe Ala Lys
145          150          155          160
Glu Val Pro Tyr Ala His Ala Asn Trp Ala Asp Ala Ala Ser His Pro
      165          170          175
Gln Asp Val Ala Ser Arg Pro Ile Trp Asp Met Gly Asn Ala Ser Leu
      180          185          190
Glu Arg Asp Leu Ser Gly Gly Trp Phe Asp Ala Gly Asp Phe Asn Lys
      195          200          205
Tyr Ser Glu Trp Thr Gly Arg Val Ile Leu Glu Leu Leu Leu Ala Tyr
      210          215          220
Gln Gly Arg Pro Asp Val Phe Thr Asp Asp Phe Gly Ile Pro Glu Ser
225          230          235          240
Gly Asn Gly Val Pro Asp Leu Leu Asp Glu Val Lys Trp Gly Met Asp
      245          250          255
Trp Leu Leu Arg Met Gln Glu Pro Ser Gly Ala Ile Leu Gly Lys Val
      260          265          270
Ser Val Thr Gly His Gln Ser Ala Ser Pro Pro Ser Thr Asp Thr His
      275          280          285
Pro Arg Tyr Tyr Gly Pro Val Ser Thr Glu Ala Thr Ala Met Ala Ala
      290          295          300
Ala Ala Phe Ala Leu Gly Ala Thr Val Phe Glu Ser Val Gly Met Ser
305          310          315          320
Asp Tyr Ala Val Thr Leu Glu Ser Ala Ala Ile Ala Ala Trp Asn Trp
      325          330          335
Thr Met Val His Pro His Val Pro Phe Asp Asn Thr Gly Phe Ala Ser

```

			340					345					350			
Val	Ser	Pro	Ser	Arg	Asn	Ala	His	Asp	Thr	Leu	Ala	Asn	Arg	Val	Met	
		355						360					365			
Ala	Ala	Ala	Met	Leu	Phe	Glu	Arg	Thr	Gly	Gly	Ala	Val	Tyr	Arg	Asp	
	370					375					380					
Phe	Phe	Asp	Val	Arg	Tyr	Leu	Asp	Met	Glu	Pro	Val	Gln	Trp	Trp	Tyr	
385					390					395					400	
Phe	Phe	Pro	Phe	Gln	Gly	Glu	Leu	Gln	Lys	Ala	Leu	Ala	His	Tyr	Thr	
				405					410						415	
Thr	Leu	Pro	Gly	Ala	Thr	Pro	Ser	Val	Ser	Ala	Asp	Ile	Arg	Asn	Arg	
			420					425						430		
Met	Ala	Ala	Ser	Ile	Asn	Gly	Gly	Glu	Phe	Leu	Gly	Ala	Trp	Asn	Asn	
		435					440					445				
Gln	Thr	Asp	Ala	Tyr	Arg	Ala	Tyr	Leu	Lys	Asp	Gln	Asp	Tyr	Thr	Trp	
	450					455					460					
Gly	Ser	Asn	Lys	Thr	Lys	Ser	Gln	Ala	Gly	Phe	Phe	Phe	Glu	Gly	Val	
465					470					475					480	
Arg	Arg	Leu	Gly	Leu	Asn	Pro	Ala	Asp	Ala	Ala	Ala	His	Arg	Asp	Ala	
				485				490						495		
Ala	Met	Gly	Tyr	Leu	His	Tyr	Leu	His	Gly	Val	Asn	Pro	Met	Gly	Met	
			500					505					510			
Val	Tyr	Leu	Ser	Asn	Met	Tyr	Ala	Ser	Gly	Ala	Asp	Arg	Ala	Ala	Asn	
		515					520					525				
Glu	Ile	Tyr	His	His	Trp	Phe	Arg	Asp	Gly	Arg	Thr	Gly	Thr	Met	Pro	
	530					535					540					
Ser	Leu	His	Ser	Thr	Val	Pro	Leu	Leu	Val	Phe	Phe	Arg	Ala	Gly	Pro	
545					550					555					560	
Asn	Ala	Gln	Ile	Gln	Arg	Lys	His	Ser	Gly	Asp	Pro	Arg	Pro	Thr	Arg	
				565					570					575		
Ala	Lys	Ser	Leu	Pro												
			580													

What is claimed is:

1. An isolated polynucleotide comprising a nucleic acid sequence with at least 85% sequence identity to SEQ ID NO: 2, wherein the polynucleotide encodes a polypeptide with cellulose hydrolyzing activity.
2. The isolated polynucleotide of claim 1 comprising a nucleic acid sequence with at least 95% sequence identity to SEQ ID NO: 2.
3. The isolated polynucleotide of claim 1, wherein the nucleic acid sequence comprises SEQ ID NO: 2.
4. An isolated polynucleotide comprising a nucleic acid sequence with at least 85% sequence identity to SEQ ID NO: 1, wherein the polynucleotide encodes a polypeptide with cellulose hydrolyzing activity.
5. The isolated polynucleotide of claim 4, comprising a nucleic acid sequence with at least 95% sequence identity to SEQ ID NO: 1.
6. The isolated polynucleotide of claim 4, wherein the nucleic acid sequence comprises SEQ ID NO: 1.
7. An isolated polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 3, wherein the polypeptide has cellulose hydrolyzing activity.
8. The isolated polynucleotide of any one of claims 1 to 7 wherein the polynucleotide is mRNA, DNA, cDNA or genomic DNA.
9. The isolated polynucleotide of claim 8, wherein said polynucleotide is an RNA molecule.
10. An expression vector comprising the polynucleotide of any one of claims 1 to 7.

11. An expression vector comprising an isolated polynucleotide of any one of claims 1 to 7, operably linked to control sequences recognized by a host cell transformed with the vector.
12. An expression vector according to claim 11 comprising a regulatory polynucleotide sequence including a promoter sequence derived from a glucose isomerase gene of *Actinoplanes*, and a signal sequence derived from a *Streptomyces* cellulase gene.
13. A host cell transformed with the vector of claim 10.
14. The host cell of claim 13, which is a prokaryotic cell.
15. The host cell of claim 13, which is a eukaryotic cell.
16. An isolated polypeptide comprising an amino acid sequence with at least 85% sequence identity to SEQ ID NO: 3, wherein said polypeptide has cellulose hydrolyzing activity.
17. The polypeptide of claim 16 comprising an amino acid sequence with at least 95% sequence identity to SEQ ID NO: 3.
18. The polypeptide of claim 16 comprising the amino acid sequence of SEQ ID NO: 3.
19. A method of producing a cellulase comprising the steps of:
 - (a) culturing the host cell according to claim 13 in a suitable culture medium under suitable conditions to produce the cellulase;
 - (b) obtaining said produced cellulase.
20. The method of claim 19 wherein the host cell is a filamentous fungi or yeast cell.

21. The method of claim 19 wherein the host cell is a bacterium.
22. The method of claim 21 wherein the bacterium is a *Streptomyces*.
23. A composition comprising the polypeptide of any one of claims 16 to 18 and a surfactant.
24. The composition according to claim 23, wherein said composition is a laundry detergent.
25. The composition according to claim 23, wherein said composition is a dish detergent.
26. A feed additive comprising the polypeptide of any one of claims 16 to 18.
27. A method of treating wood pulp comprising contacting said wood pulp with a polypeptide according to any one of claims 16 to 18.
28. A method of converting biomass to sugars comprising contacting said biomass with a polypeptide according to any one of claims 16 to 18.
29. The method of claim 28 further comprising the generation of high fructose corn-syrup
30. A method of producing ethanol, said method comprising the steps of:
 - (a) contacting a biomass composition with an enzymatic composition comprising the polypeptide of any one of claims 16 to 18 to yield a sugar solution;
 - (b) adding to the sugar solution a fermentative microorganism; and
 - (c) culturing the fermentative microorganism under conditions sufficient to produce ethanol.

1/4

Figure 1A Nucleotide sequence of inserted environmental DNA (029cel)
SEQ ID NO:1

ATCAACACGC	TGGAAAGTAA	TTTCAAGGGT	AAGGCCATCG	GTTGCCGCCG	50
GGGTAGAAAT	GTGCGGTTGG	ATTTCGTTGA	GCGGCGTCGC	CGGCGTTCCA	100
CCGAGGGCAT	AGCGCAGCAG	GTTGGCGATG	CCACCGGTGA	GGCCTTCGGG	150
GCCGCCTACG	ATGTTGTGCT	CAGCCGCCCA	TGCGATGTAG	CCGTCCGGCT	200
CGGGTTCGCT	CGCGGGGGTG	AAGAAGACAA	TGTCGTGAG	ATAAAGGTTG	250
CCGCTTCCGC	TCTCAACGCC	GCCGAGGTTG	AATTGGATTT	CGCAAATTCT	300
CGTTAGGTCC	AGCACGGAAT	CGCCGACGAG	GTCGGCTATG	GGAATCTGAA	350
TGCGCCCAT	GGGTTGGGTA	CGCGGAAGGG	ACACGTAGGG	ACCCACTTTG	400
TCATTGGGCG	AGACGAGCCG	GACAAAGATT	TGGTGCGCCG	CCTGCGAGGG	450
GCCTTGGAGG	GCGAGAGAAA	GGTACGTGAG	GGCGCTGATG	TCGTGCGTGG	500
GACCGTCTCC	CCAGTTGTCG	AGATTGAGCC	CAAATCCGGC	CCACCATCCG	550
GCGATAGTGT	AGCTCCAATG	GTAGTGACGC	TCACCCTCGA	AGCCGCCGCT	600
GGAGAGTTCC	TGCAAGCCGT	CGCCCCAAAT	GCCCGTGATG	AGCGTTGCCT	650
CGTCACGGTA	GATCACAAGT	TCGGCGGCGG	GTGCCGGGGG	AAGATCGCCT	700
TGAGTGATCA	CGAGAGTGGC	GGTGGCGCTG	CCTTCGTGAT	TAGGGTCGGT	750
AATGGTGGCG	ACGACCGTGT	AGCTACCGGG	CCCCACTGGC	GCATGGGTGG	800
AACCGTTGTA	GGTAAAGGAG	ACGTCAAGCC	CCACGGGATG	GGTCTCGGCA	850
AGAGCGGCCT	TGGGGGTGCC	GTCGAAAACG	TGTTCCAAAT	TGGAGAGCGT	900
GATGGTGGCG	GGTGCCTTGA	GCACAGTCAC	AGAAACAGTG	GATTGCACGG	950
GATCGTGCGC	TGCCGTGTCT	GCAGGTGTGA	AGACCACGCT	GTAAAAACGG	1000
GTTCCGGCGG	ACGGTGCAAG	GCCGGACAGG	ACAAAGGCAA	AGTCGCCGGG	1050
GACGGCGGCT	ACTCCGCCGC	TCAGGCCGGC	CTCCGCAAGG	GTTTGCCCGA	1100
AGGTGATGGG	TGCGGCTGTG	GGCCACATCT	CCACAAGGCC	GGTGTCCCCC	1150
TCGTCACGCA	CCGGCATGAG	GGCGGAGAGG	AGATGAATGT	AACTGGCTTG	1200
GTAATTGATG	TCGGGCTCGG	TGATTTCCCA	TGAGTTCTCC	GGCCAAAAC	1250
CATTCCAATC	AAGGTAGGCT	TTTTGCACGG	GTTGGTCTCG	GATCGCCTGA	1300
ATGCTTCCGC	TGTATTTGGG	CATTGGGACC	CGCCCGAAAG	AAAACCAGGA	1350
GCGGGACCGT	AGAGTGAAGT	GAGGGCATTG	TCCCAGTCCG	GCCATCGCGG	1400
AACCAATGGT	GGTAGATTTT	ATTGGCTGCA	CGGTCAGCGC	CGCTGGCATA	1450
CATGTTGCTA	AGATAGACCA	TGCCCATTGG	GTTCACTCCG	TGGAGATAGT	1500
GCAGGTAGCC	CATCGCGGCA	TCGCGATGCG	CGGCCGCGTC	GGCGGGGTTG	1550
AGCCCAAGCC	TCCGTACCCC	CTCGAAGAAA	AAGCCAGCCT	GAGACTTTGT	1600
TTTGTTCGAG	CCCCACGTGT	AATCCTGATC	CTTCAGGTAG	GCGCGGTAGG	1650
CGTCGGTCTG	GTTATTCCAT	GCACCGAGAA	ACTCCCCACC	GTTTATAGAA	1700
GCCGCCATCC	GGTTGCGGAT	GTGCGCAGAG	ACGCTAGGCG	TCGCTCCCGG	1750
GAGGGTCGTG	TAGTGGGCGA	GAGCTTTTTG	TAGCTCACCT	TGAAAGGGGA	1800
AGAAATACCA	CCACTGCACG	GGCTCCATAT	CGAGATAGCG	CACATCGAAG	1850
AAATCGCGAT	AGACCGCACC	GCCCGTGCGC	TCGAAGAGCA	TGGCGGCGGC	1900
CATCACACGG	TTGGCTAGCG	TATCGTGGGC	ATTGCGCGAG	GGGCTCACGG	1950
AAGCAAATCC	GGTGTGTCG	AAAGGCACAT	GAGGATGGAC	CATGGTCCAA	2000
TTCCATGCGG	CGATGGCAGC	GGATTGAGG	GTGACGGCAT	AATCGCTCAT	2050
GCCTACGCTC	TCAAAGACAG	TCGCCCCGAG	GGCGAAAGCG	GCGGCAGCCA	2100
TGGCAGTGCC	CTCGGTCGAG	ACGGGGCCGT	AGTAACGCGG	ATGGGTGTCG	2150
GTGCTCGGCG	GGCTGGCGCT	CTGGTGCCCC	GTCACGGAAA	CTTTCCCGAG	2200
AATAGCCCCG	CTCGGCTCCT	GCATGCGTAA	GAGCCAGTCC	ATTCCCCATT	2250
TGACTTCGTC	AAGCAGGTCG	GGGACACCGT	TGCCGGATTC	CGGGATGCCA	2300
AAATCATCGG	TAAAGACGTC	AGGCCGCCCT	TGATAGGCAA	GGAGCAGCTC	2350
CAGGATGACG	CGCCCCGTCC	ACTCGCTGTA	CTTGTTGAAA	TCGCCCCGAT	2400

2/4

Figure 1B

CGAACCAACC	GCCGCTGAGA	TCGCGCTCCA	AGGAGGCATT	CCCCATATCC	2450
CAGATGGGGC	GGCTGGCGAC	GTCCTGCGGG	TGAGAAGCGG	CATCGGCCCA	2500
GTTCGCGTGG	GCGTAGGGCA	CCTCCTTGGC	AAACCCGGAG	CGCTGATAGA	2550
AGAACATGCG	CACGGCCTCG	CGCAGGACAA	CATCGTAAAC	ATCCGCGCCA	2600
ATGGCGAAAC	TATCGGAATG	AGTGTTGTTG	GCAGGATCGT	GGATGCGGTA	2650
GTGGCCGGGC	TCGGCAACTA	CCGTAAAATC	AAACCACCAC	ACGCGGTCTC	2700
CCGATTGAAT	ATGGATGGCG	CCGCCGTTCC	ACGGGACCGG	TGAGCCGGAG	2750
AAAACCACGA	CGCCATCGTT	CACGCGACGG	ACCTCCAGCG	TTGCGCCGGG	2800
GCTGTAGCTC	TCGGCGCTGT	TCCAGCCAAT	CTGCGGGTCG	GCGATCACCG	2850
CCACCTTGGT	GGCATCGGCG	GGGTAACCGA	ATTGGTCGAT	GCGGATTTTA	2900
TCGGTGTGGG	TGGAGGCGAC	GAGGGCGGAG	CTGCCCATGA	GCAGCAAGAA	2950
AAAGCCCGCT	GTCGGCCCGA	TACCAAAAAA	ACGAATAGGG	AGAGAAAAAT	3000
TCATAGCAGG	ATGTGGATAC	GGAAAGGGGG	AAAACGGTGC	AAAGACCCAA	3050
GCCCAACGCT	TGGCGAAAAC	TGGATGGTTG	GTTTATCAAG	AAAAGCGCTT	3100
TTGAGCCAAA	AGCTGCGGGC	AATCCTTATT	GCGTTTCACA	ATATTTTCAC	3150
ATCGTCGGCG	GCACGACTTT	TCGATGGGCG	ACTTGACAGC	GTATTCTCTC	3200
AGGCGCGAGG	CTGCAAACCT	TATGAAAAAA	GGCCCGCGCA	GCGATCTGTC	3250
CCCGGTCAAA	ATCCAGTCAA	GGTTTGTTCA	AGGGTTTGAG	GTCTGATAGA	3300
GGCACAGTCG	AGCCATCAGC	AGTCGCATTG	AGTAGGGTTG	TTGGAGAAAG	3350
TGTGCAAATG	ACCGCTGCCG	AAGGAACTGT	GGAGACAAAA	AGCATATTTT	3400
CCTCGCCAAG					3410

Figure 2
The nucleotide sequence of 029cel ORF

ATGAATTTTT	CTCTCCCTAT	TCGTTTTTTT	GGTATCGGGC	CGACAGCGGG	50
CTTTTCTTG	CTGCTCATGG	GCAGCTCCGC	CCTCGTCGCC	TCCACCCACA	100
CCGATAAAAT	CCGCATCGAC	CAATTCGGTT	ACCCCGCCGA	TGCCACCAAG	150
GTGGCGGTGA	TCGCCGACCC	GCAGATTGGC	TGGAACAGCG	CCGAGAGCTA	200
CAGCCCCGGC	GCAACGCTGG	AGGTCCGTCG	CGTGAACGAT	GGCGTCGTGG	250
TTTTCTCCGG	CTCACCGGTC	CCGTGGAACG	GCGGCGCCAT	CCATATTCAA	300
TCGGGAGACC	GCGTGTGGTG	GTTTGATTTT	ACGGTAGTTG	CCGAGCCCGG	350
CCACTACCGC	ATCCACGATC	CTGCCAACAA	CACTCATTC	GATAGTTTCG	400
CCATTGGCGC	GGATGTTTAC	GATGTTGTCC	TGCGCGAGGC	CGTGCGCATG	450
TTCTTCTATC	AGCGCTCCGG	GTTTGCCAAG	GAGGTGCCCT	ACGCCACGC	500
GAACTGGGCC	GATGCCGCTT	CTCACCCGCA	GGACGTCGCC	AGCCGCCCCA	550
TCTGGGATAT	GGGGAATGCC	TCCTTGAGC	GCGATCTCAG	CGGCGGTTGG	600
TTCGATGCGG	GCGATTTCAA	CAAGTACAGC	GAGTGGACGG	GGCGCGTCAT	650
CCTGGAGCTG	CTCCTTGCCT	ATCAAGGGCG	GCCTGACGTC	TTTACCGATG	700
ATTTTGGCAT	CCCGGAATCC	GGCAACGGTG	TCCCCGACCT	GCTTGACGAA	750
GTCAAATGGG	GAATGGACTG	GCTCTTACGC	ATGCAGGAGC	CGAGCGGGGC	800
TATTCTCGGG	AAAGTTTCCG	TGACGGGGCA	CCAGAGCGCC	AGCCCGCCGA	850
GCACCGACAC	CCATCCGCGT	TACTACGGCC	CCGTCTCGAC	CGAGGCCACT	900
GCCATGGCTG	CCGCCGCTTT	CGCCCTCGGG	GCGACTGTCT	TTGAGAGCGT	950
AGGCATGAGC	GATTATGCCG	TCACCCTCGA	ATCCGCTGCC	ATCGCCGCAT	1000
GGAATTGGAC	CATGGTCCAT	CCTCATGTGC	CTTTCGACAA	CACCGGATTT	1050
GCTTCCGTGA	GCCCCTCGCG	CAATGCCCAC	GATACGCTAG	CCAACCGTGT	1100
GATGGCCGCC	GCCATGCTCT	TCGAGCGCAC	GGGCGGTGCG	GTCTATCGCG	1150
ATTTCTTCGA	TGTGCGCTAT	CTCGATATGG	AGCCCGTGCA	GTGGTGGTAT	1200
TTCTTCCCCT	TTCAAGGTGA	GCTACAAAAA	GCTCTCGCCC	ACTACACGAC	1250
CCTCCCGGGA	GCGACGCCTA	GCGTCTCTGC	CGACATCCGC	AACCGGATGG	1300
CGGCTTCTAT	AAACGGTGGG	GAGTTTCTCG	GTGCATGGAA	TAACCAGACC	1350
GACGCCTACC	GCGCCTACCT	GAAGGATCAG	GATTACACGT	GGGGCTCGAA	1400
CAAAACAAAG	TCTCAGGCTG	GCTTTTTTCTT	CGAGGGGGTA	CGGAGGCTTG	1450
GGCTCAACCC	CGCCGACGCG	GCCGCGCATC	GCGATGCCGC	GATGGGCTAC	1500
CTGCACTATC	TCCACGGAGT	GAACCCAATG	GGCATGGTCT	ATCTTAGCAA	1550
CATGTATGCC	AGCGGCGCTG	ACCGTGCAGC	CAATGAAATC	TACCACCATT	1600
GGTTCCGCGA	TGGCCGGA	GGGACAATGC	CCTCACTTCA	CTCTACGGTC	1650
CCGCTCCTGG	TTTTCTTTTCG	GGCGGGTCCC	AATGCCCAA	TACAGCGGAA	1700
GCATTCAGGC	GATCCGAGAC	CAACCCGTGC	AAAAGCCTA	CCTTGA	1746

4/4

Figure 3**The translated 029cel protein composed of 581 amino acids**

MNFSLPPIRFF	GIGPTAGFFL	LLMGSSALVA	STHTDKIRID	QFGYPADATK	50
VAVIADPQIG	WNSAESYSPG	ATLEVRRVND	GVVVFSGSPV	PWNGGAIHIQ	100
SGDRVWWFDF	TVVAEPGHYR	IHDPANNTHS	DSFAIGADVY	DVVLREAVRM	150
FFYQRSQFAK	EVPIYAHANWA	DAASHPQDVA	SRPIWDMGNA	SLERDLSSGGW	200
FDAGDFNKYS	EWTGRVILEL	LLAYQGRPDV	FTDDEFGIPES	GNGVPDLLDE	250
VKWGMDWLLR	MQEPSGAILG	KVSVTGHQSA	SPPSTDTHPR	YYGPVSTEAT	300
AMAAAAFALG	ATVFESVGMS	DYAVTLESAA	IAAWNWTMVH	PHVPFDNTGF	350
ASVSPSRNAH	DTLANRVMAA	AMLFERTGGA	VYRDFFDVRY	LDMEPVQWWY	400
FFPFQGELOK	ALAHYTTLPG	ATPSVSADIR	NRMAASINGG	EFLGAWNNQT	450
DAYRAYLKDQ	DYTWGSNKTK	SQAGFFFEGV	RRLGLNPADA	AAHRDAAMGY	500
LHYLHGVNPM	GMVYLSNMYA	SGADRAANEI	YHHWFRDGRT	GTMPSLHSTV	550
PLLVFFRAGP	NAQIQRKHSG	DPRPTRAKSL	P		581

The translated 029cel protein composed of 581 amino acids

MNFSLPPIRFF	GIGPTAGFFL	LLMGSSALVA	STHTDKIRID	QFGYPADATK	50
VAVIADPQIG	WNSAESYSPG	ATLEVRRVND	GVVVFSGSPV	PWNGGAIHIQ	100
SGDRVWWFDF	TVVAEPGHYR	IHDPANNTHS	DSFAIGADVY	DVVLREAVRM	150
FFYQRSGFAK	EVPYAHANWA	DAASHPQDVA	SRPIWDMGNA	SLERDLSGGW	200
FDAGDENKYS	EWTGRVILEL	LLAYQGRPDV	FTDDEFGIPES	GNGVPDLLDE	250
VKWGMDWLLR	MQEPSGAILG	KVSVTGHQSA	SPPSTDTHPR	YYGPVSTEAT	300
AMAAAAAFALG	ATVFESVGMS	DYAVTLESAA	IAAWNWTMVH	PHVPFDNTGF	350
ASVSPSRNAH	DTLANRVMAA	AMLFERTGGA	VYRDFFDVRY	LDMEPVQWWY	400
FFPFQGELOK	ALAHYTTLPG	ATPSVSADIR	NRMAASINGG	EFLGAWNNTQ	450
DAYRAYLKDQ	DYTWGSNKTK	SQAGFFFEGV	RRLGLNPADA	AAHRDAAMGY	500
LHYLHGVNPM	GMVYLSNMYA	SGADRAANEI	YHHWFRDGRT	GTMPSLHSTV	550
PLLVFFRAGP	NAQIQRKHSG	DPRPTRAKSL	P		581